

Can physics deliver another biological revolution?

Cultural, institutional, conceptual and linguistic barriers are being overcome as physicists and biologists recognize the scientific stimulus they can gain from each other. The United States is showing the way.

What can physics do for molecular biology? This may seem an odd question in an era when biomedical science is lavishly funded, attracts widespread public interest and shows no sign of slackening its pace of discovery. Surely molecular biology, of all the sciences, should be able to take care of itself.

Yet the question is being asked — and answered — by increasing numbers of biologists and physicists, and by those who employ and fund them. And the conspicuous success of molecular biology is one of the main drivers of the trend. The genetic engineer's tool kit has been enormously powerful, but only recently have techniques become available that can provide high volumes of quantitative biological information. For example, it is now becoming routine to monitor the expression of up to 10,000 genes at a time, using DNA microarray technology. From genome sequencing to protein structure determination, biologists are increasingly thinking about how to manage 'floods of data' — and realizing that physical scientists have been coping with this problem for decades.

Beyond simply 'managing' the data — for which the tools of computer science and informatics are proving essential — biologists are finding that they also need new ways of thinking about the data. It is only a small exaggeration to say that the main method of analysis in molecular biology has been the cartoon representation of networks, pathways and complexes; indeed, superb papers have been written for the purpose of adding a single arrow to an existing cartoon. But to really understand the biochemical network thus represented, one needs to have numbers attached to the arrows, and equations relating the numbers. The paper by Alon *et al.* on pages 168–171 of this issue shows how such a model — combined with the powerful methods of genetic engineering — has led to a deeper understanding of bacterial chemotaxis.

So what's new?

If this were just a story of biologists co-opting tools from the physical sciences, nothing much would be new. Nor is the story one of physicists simply 'doing physics' in biological systems — for example, studying the properties of DNA as a model for other polymers. What's new is that many physicists — not just a few isolated pioneers — are getting excited by the challenge of tackling important questions in biology, using the tools, both physical and mental, of physics.

One should not underestimate the extent of this challenge. History records many sad cases of physicists enthusiastically attacking biological problems that they could solve, only to find that biologists considered these problems uninteresting. Physicists are also addicted to simplification — a habit viewed with suspicion by biologists, who know from hard-won experience how much small details can matter. Thus, the experimental physicist will find that it is no trivial matter to control all but one of the variables in a biological system. Meanwhile, the theorist, used to the productive interplay of theory and experiment that seems essential for progress in physics, will find that most molecular biologists have little time for mathematical theory,

which has played no significant role in their field's great advances.

Then there is the language barrier, and the sheer number of facts and labels to be assimilated before a physicist can have a productive conversation with a biologist collaborator. The biologist, in turn, will probably have had very little quantitative training — increasing the difficulties of finding common ground. The determined interdisciplinarians who surmount all of these hurdles may then find institutional obstacles in their way. Physics departments may be loath to hire physicists who work on biological problems, for fear that they will spend all their time in the biology department. Cries of 'That's not really physics!' and 'But can he teach quantum mechanics?' have been known to resound in department meeting rooms when biophysicists have been considered for faculty positions.

The physics–biology agenda

The good news is that there seems to be a considerable will to overcome these and other institutional obstacles. Several universities in the United States have announced plans to put physical scientists and biologists together in interdisciplinary institutes, while they retain their appointments in existing departments (*Nature* 397, 3; 1999). Physics and biology departments are learning to make joint appointments, sometimes with one department providing the post and the other providing laboratory space. And one physicist has spoken of the thrill of installing the first real biology laboratory in a physics building — "although there are no mice or monkeys running around yet".

In the United States, government and private funding agencies are also promoting the physics–biology agenda. The National Institutes of Health (NIH) has several initiatives designed to bring physical scientists into biology, and a National Science Foundation programme funds mathematical and physical scientists and engineers who want to collaborate with scientists at NIH. The Sloan Foundation (with the US Department of Energy) funds fellowships for physical scientists in computational molecular biology, and the Burroughs Wellcome Fund gives five-year grants to institutions that provide graduate and postdoctoral training at the physics–biology interface.

It should be stressed that, appropriately, all of this activity amounts to a small part of physics and biology. No one is suggesting that a large fraction of condensed-matter physicists should abandon their inanimate pursuits to get their hands wet in biology labs; and in biology there is still plenty of mileage left in developing cartoons. But it may be well to recall that today's molecular biology had its origins half a century ago in the work of a small band of (mostly) physicists. And although Max Delbrück's model for the molecular origin of mutations, popularized in Erwin Schrödinger's classic book *What is Life?*, proved to be wrong, these physicists introduced ways of approaching biological problems that stimulated a revolution in biology. Today, with physicists who can manipulate single molecules in the laboratory, and simulate and quantitatively analyse complex systems, who can say what might not be possible? □



US biologists propose launch of electronic preprint archive

[PARIS] An electronic challenge to traditional biomedical science journals is being planned by a group of US scientists. They propose to create an e-print archive similar to that established for physics at the Los Alamos National Laboratory in New Mexico by Paul Ginsparg in 1991.

The initiative is led by Patrick Brown, a researcher at the Howard Hughes Medical Institute at Stanford University School of Medicine, and David Lipman, director of the US National Council for Biotechnology Information (NCBI), which operates the PubMed and GenBank databases.

Under preliminary proposals, the life science e-print archives would be established at a single website. Papers could be automatically posted, archived and distributed, with authors retaining copyright over their work. Articles would either be posted without refereeing, or peer reviewed by third-party professional societies or electronic journals, and labelled as such. Signed commentaries would also be posted along with articles.

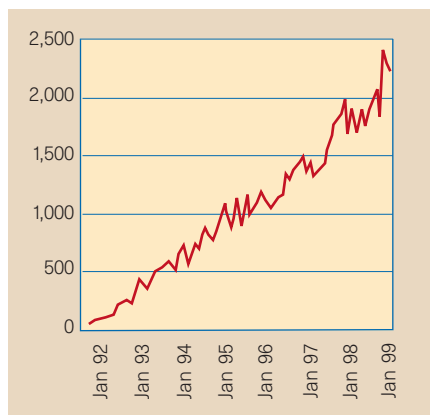
E-print archives were first established in high-energy physics as an electronic alternative to distributing preprints. But by automating the process, Ginsparg created a large free electronic journal for papers that have not been peer reviewed (see graph). This year the archive will 'publish' some 25,000 papers in astrophysics, quantum physics, condensed matter theory and computer science.

The Los Alamos archives are the primary means of first publication in many areas of physics, although many papers are later published in journals. E-print archives for cognitive science have been established by Steven Harnard at the University of Southampton, and a computer research repository for the computer sciences was recently launched by Joseph Halpern and Carl Lagoze at Cornell University, New York state.

Lipman said last week that it was too early to comment on the plans in detail. But under one proposal being discussed, the archives would benefit from direct indexing and linking via NCBI's PubMed database.

Brown says this would guarantee wide exposure, as PubMed is one of the most widely used Internet services in the biomedical sciences. It offers access to 9 million bibliographic citations from 4,000 journals in Medline, and publishes abstract information provided by publishers before it has been indexed and made available on Medline.

PubMed has hyperlinks to full text in more than 300 journals, and links molecular biology citations to protein and nucleotide sequence databases, and from one Medline



Trend setter: monthly submissions to the physics archive at Los Alamos National Laboratory have risen sharply since its launch in August 1991.

record to other related entries. The number of daily users has climbed from 35,000 to 70,000 over the past year, and PubMed now handles 1.5 million transactions daily.

Access to the proposed e-print archives would be free. The protagonists are said to be seeking funding for operating costs from the Howard Hughes Medical Institute and the US National Institutes of Health.

The initiative is likely to be welcomed in the research community. Fotis Kafatos, director of the European Molecular Biology Laboratory in Heidelberg, Germany, says: "I believe strongly that more open communication could benefit all the life sciences."

Kafatos says he would not like to see the established journals system dismantled. But he believes that an e-print archive would undermine the "scoop mentality" of some

journals towards the acquisition of papers, and prompt a shift towards publishers providing greater editorial value.

The proliferation of specialized journals has compartmentalized knowledge in an artificial way, argues Kafatos. "Obviously the content of journals like *Nature* and *Science* appeals to a wide audience. But as a whole, biology is rapidly moving in the direction of dismantling barriers, and you need new ways to cut across disciplines." Electronics provides means for "more flexible organization of knowledge," he says.

Brown intends to invite leaders in the biomedical community to a workshop to discuss the proposals and encourage scientists to publish in the archive; biologists, unlike physicists, have no strong tradition of distributing e-prints.

The main obstacle facing the archive is the fear that publishing an e-print will preclude later publication in a peer-reviewed journal. Several journals (including *Nature*) have relatively liberal policies stating that publishing e-prints does not preclude publication, and others are expected to follow.

But some journals maintain that publishing on the Internet constitutes prior publication. Brown wrote to life-science journals this week asking them to publish "an explicit policy statement that distribution of a preprint, by means of a public electronic preprint server or Internet site, will not influence the decision of your journal to publish a paper". The letter was signed by 30 scientists including James Watson, Arthur Kornberg and Daniel Botstein.

Declan Butler

Next week: Briefing on electronic journals

Indian scientist sues over spying charge

[NEW DELHI] A rocket scientist at the Indian Space Research Organization (ISRO) in Bangalore who was falsely charged with spying five years ago has issued a \$250,000 suit against the government of Kerala state and the central government in New Delhi.

In a suit filed in a court in Thiruvananthapuram on 2 January, Nambi Narayanan, currently director of the ISRO's advanced technology and planning department, is demanding compensation for the "mental and physical agony" he claims to have suffered at the hands of the Kerala state police and officials of the central government's intelligence bureau.

Narayanan seeks damages on ten counts, including false arrest, illegal detention for 50 days, mental and physical torture, loss of

reputation, distress to his family and loss of a promotion. The court has not set a date for the hearing.

In 1994, Narayanan, then deputy director of ISRO's liquid propulsion systems centre, and three others were charged with leaking information on ISRO's cryogenic engine project to an unidentified foreign power through two women alleged to be agents of Maldives (see *Nature* 372, 491; 1994).

The case was dismissed as groundless by India's Supreme Court in April 1998 (see *Nature* 393, 6; 1998), and the two women were released in October. Narayanan says he is suing the government because the officials who "concocted the case" and tortured him had refused to admit their guilt and had not been punished.

K. S. Jayaraman

Future of German neutron source hangs in the balance

[MUNICH] The planned opening in 2001 of a controversial 20-MW research reactor being built by the Technical University of Munich at Garching could depend on the outcome of discussions being held this week by Germany's new coalition government on its nuclear policy.

In particular, the discussions will address a proposal from Green environment minister Jürgen Trittin to bar the licensing of new nuclear reactors — including research reactors used as a source of neutrons for medical and material sciences — of more than 1 MW thermal power.

The reactor concerned, FRM II, has been strongly opposed by environmental groups because it has been designed to burn bomb-grade, highly enriched uranium (HEU), in defiance of a 1978 international anti-nuclear proliferation agreement that all reactors should convert, if technically possible, to low enriched uranium.

An amendment to this agreement requires the United States to take back HEU waste from reactors agreeing to convert (see *Nature* 370, 90; 1994). The Technical University overcame opposition to win building licences, and the DM770 million (US\$456 million) building operation is well under way. But the reactor requires an operation licence, for which the university must apply next year.

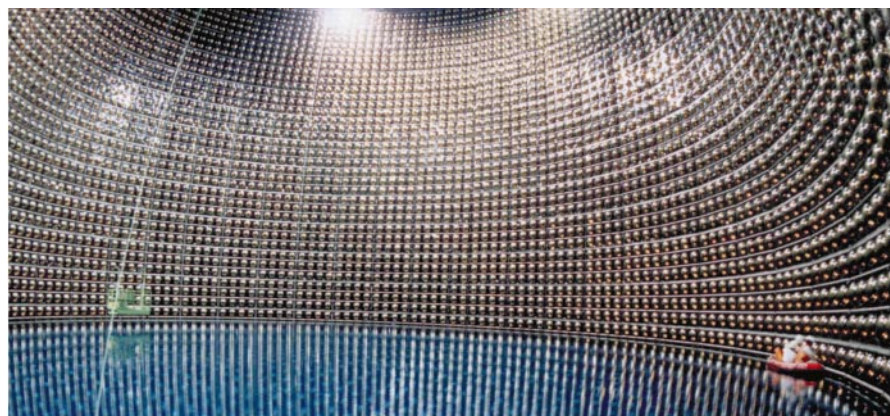
Trittin's proposal, leaked to the press last week, also includes plans to bar the transport or reprocessing of nuclear waste in Germany. He argues that waste should be stored above ground at interim storage sites close to the reactors themselves. But this proposal has concerned research reactor operators.

The Hahn Meitner Institute in Berlin, for example, whose reactor is converting to LEU, was recently forced to send a consignment of spent fuel rods to the UK reprocessing facility in Dounreay, Scotland, because the US agreement to take back spent fuel was in temporary abeyance and its own limited storage facilities were dangerously overloaded (see *Nature* 377, 470; 1995).

If the shipped fuel has to be returned, it could cause a major dilemma. A spokesman for the institute says plans to build a new storage facility would take years to gain approval. Moreover, he says, the implications are unclear for spent fuel recently shipped to the United States, as US officials have not decided whether it should be stored or reprocessed.

A spokeswoman for the German research ministry says this week's talks are intended to clarify such issues. Meanwhile, the Technical University has mounted a vigorous press campaign opposing any decision to prevent its reactor from opening.

Alison Abbott



The Super-Kamiokande neutrino detector: results will throw light on cosmological theories.

Japanese centre will study the origins of the Universe

[TOKYO] The University of Tokyo has announced plans to set up a research centre to study the 'Big Bang' model of the origins of the Universe from both observational and theoretical perspectives.

One goal of the centre could be to explore a new and more flexible theory for solving the problems posed by the presence of free parameters in this model which has long restricted its predictive power.

Researchers from the university say their interest in the theory has been strengthened by recent findings from distant supernova explosions which suggest that the Universe entered a period of intense inflation shortly after the Big Bang, as advocated by the theory of 'cosmological inflation'.

The centre will be set up in April within the university's department of science. It will replace the Research Centre for the Early Universe (RESCEU), a five-year project funded by the Ministry of Education, Science, Sports and Culture (Monbusho) to study the origin and evolution of the Universe.

Although RESCEU, which began in 1995, should not end until March next year, its replacement by the new research centre will be a year earlier than scheduled. This is to allow the earliest possible start to the new project, which will merge five research divisions from RESCEU with three new divisions, including one set up specifically for overseas researchers.

The centre, which will conduct research in areas such as the very early Universe, dark matter, cosmic rays and the formation and evolution of galaxies, will be part of Monbusho's 'centre of excellence' programme, which supports research organizations carrying out 'highly creative scientific research'.

"An early start will allow us to take a step towards our aim to create an international base for research on the Big Bang model," says Katsuhiko Sato, professor of theoretic-

cal astrophysics and the director of the new centre.

"We feel this is the time to give substance to the theoretical model by applying results from areas such as nuclear and particle physics as well as data obtained from the latest observations."

According to Sato, the centre will consist of two sections: theories of the early Universe, which will take a 'top down' approach based on the theory of cosmological inflation developed by Guth, Linde and Steinhardt in the early 1980s, which proposes a very rapid expansion of the Universe shortly after the Big Bang; and a 'bottom up' approach of creating models using observational data and results obtained from experiments in particle physics.

It is hoped that the observations of neutrinos from supernova explosions, such as those by researchers from Tokyo University's Institute for Cosmic Ray Research using the Super-Kamiokande neutrino detector, will contribute to the systematic study of particles needed to understand the early Universe.

Studies of the early Universe, including research on its birth and on cosmological phase transitions, primordial nucleosynthesis, and baryogenesis (the production of protons and neutrons), based on general relativity and particle physics, will use observational data obtained from national sources, such as Monbusho's Institute of Space and Astronomical Sciences, as well as those from international sources such as the Sloan Digital Sky Survey.

"The importance of the project lies in the fact that our research will be driven by both theory and observational data. This will allow quantitative comparisons between results from observations and those from theories, and will also help to integrate research results from different fields," says Sato.

Asako Saegusa

China signs up for participation in EU's Framework

[MUNICH] The European Commission has signed a cooperation agreement on science and technology with the Republic of China that will allow Chinese scientists to participate in its fifth Framework programme of research, which was approved by the European Council of Ministers late last month.

The agreement gives European scientists the reciprocal right to participate in Chinese research programmes. The agreement involves no exchange of money — scientists on both sides will have to pay their own way but will have access to common results.

This is the first time the commission has signed such an agreement with a developing economy, but it will not be the last. Negotiations are continuing with Argentina and India, and informal discussions are taking place with other countries.

The European Molecular Biology Organization (EMBO) last year set up the European Biotechnology Node for Interactions with China (<http://www.ebnc.org>) to help build contacts between European and Chinese scientists, and promote their participation in the Framework programme.

Frank Gannon, executive director of EMBO, explains: "The challenge to European scientists to find their way to the fifth Framework programme is enormous. The challenge to our colleagues in China is nearly unimaginable."

Alison Abbott

Brazil brings in reforms of science funding agency

[SÃO PAULO] Brazil's science funding agency is being shaken up as part of broad government changes introduced to coincide with the second term of the country's president, Fernando Henrique Cardoso.

The new minister of science, Luiz Carlos Bresser Pereira, will temporarily head the National Council for Scientific and Technological Development (CNPq). And the administration of the council is being brought closer in line with similar funding agencies abroad.

Cardoso had considered merging the Ministry of Science and Technology with the larger Ministry of Education (see *Nature* 396, 503; 1998). But he decided to keep them separate, and make the changes inside each.

Bresser Pereira replaced José Israel Vargas, who had headed the science ministry since the government of Cardoso's predecessor, Itamar Franco. José G. Tundisi, a former president of CNPq, is to head the committee responsible for the council's reorganization.

Fernando Reinach, a molecular biologist at the University of São Paulo and a campaigner for reforms, who had been asked to review the new structure of the CNPq, has been appointed by Bresser Pereira as his secretary for science policy (see *Nature* 392, 647–648; 1998).

The independence from political interference of the CNPq, which was created in

the late 1940s, has always been prized by Brazilian scientists. But when the science ministry was created in 1985, it took on responsibility for the agency, creating a source of friction.

The ministry, for example, had three top officials — one for scientific development, one for technological development, and a third for information technology — each of whom was involved in science policy decisions.

Under the new reforms, these will be replaced by three new CNPq vice-presidencies, each working as a separate funding council: one for biomedical research, another for humanities, and another for physical sciences and engineering. According to Reinach, each of these posts will legally have to be filled by a working scientist.

CNPq's previous responsibility for a number of scientific institutes will be taken over by the ministry. "Basically, we are increasing the participation of scientists in the ministry," says Reinach.

Several scientists will serve on the committee that will draw up the details of the reform, including Moacyr Krieger, president of the Brazilian Academy of Sciences. Others who have discussed the reform with the minister include biologist Sérgio Ferreira, president of the Brazilian Society for the Development of Science.

Ricardo Bonalume Neto

Energy department revises terms of Venter deal after complaints

[WASHINGTON] Leaders of the publicly funded Human Genome Project have rejected a proposed memorandum of understanding between one of the project's participants, the US Department of Energy (DoE), and Celera, the corporation established by geneticist Craig Venter in a rival bid to sequence the genome.

There had been concern that clauses in the proposed agreement would have been in conflict with the desire for open access to sequence data. But the leaders of the government project still hope that the DoE can reach an agreement with Celera before the corporation starts its sequencing work this spring (see *Nature* 393, 101; 1998).

The National Human Genome Research Institute (NHGRI) at the US National Institutes of Health (NIH) and the UK Wellcome Trust — the two largest participants in the international genome project — last month scotched the draft agreement under which Celera and the DoE would have exchanged sequence information on the three chromosomes of

the genome that are being sequenced by the DoE.

Francis Collins, director of the NHGRI, says his institute and the trust were "uncomfortable" with the draft because it would have required DoE scientists to work with data from Celera that was not publicly available, and it involved the government agency in providing a service to a single, profit-making corporation. The draft would have been "in violation of the international consensus to have sequence data released immediately" into the public domain.

Collins and Michael Morgan, a senior official responsible for the genome project at the Wellcome Trust, asked Ari Patrinos, head of biological and environmental research at DoE, to come up with a new agreement.

Patrinos now concedes that the draft would have "given some scientists an early peek" at genome data, and that this was "inappropriate and incorrect". Patrinos expects that a new agreement will be negotiated.

Committees of the US Congress have pressed the NIH and DoE to collaborate with Venter's effort, but differing approaches to the release of data will make this easier said than done. The government project requires all sequence material to be published immediately. Celera wants to keep its data secret for three months — scouring it for information that can be patented — before sharing it with the outside world.

Collins says he is unsure how the agreement can be made to work, but says that "no data will be exchanged that is not publicly available".

Paul Gilman, head of policy at Celera, says the corporation will be receptive to any new proposal from the public agencies, but that it "hadn't heard a word from them" since the draft was withdrawn in December.

Collins strongly denied a report in *Time* magazine that "sour grapes" at the NIH was behind the rejection of the draft. Venter has a history of friction with the NIH, which he left in 1992 after an explosive public argument over gene patents.

Colin Macilwain

Gene therapy pushes on, despite doubts

[WASHINGTON] A conference on *in utero* gene therapy organized by the US National Institutes of Health (NIH) concluded last week that many significant scientific and ethical questions must be answered before the practice can or should be attempted in humans.

But despite the obstacles — and determined opposition from activists who attended — senior NIH officials said they would continue to revise their guidelines on gene therapy to address eventual *in utero* applications. The revisions could be published as soon as June.

The Food and Drug Administration (FDA), which holds the final power to

approve *in utero* protocols, also appeared keen to push ahead. "We need to make sure that whatever we do, the very best technology is used," Philip Noguchi, director of the FDA's Division of Cellular and Gene Therapies, told the meeting. "And the only way to get there is to really force the issue."

The technique's immaturity was clear at the meeting. Experts pointed to a dearth of animal data, and to human risks including inadvertent genetic modification of fetal sperm and eggs, or even of the rapidly proliferating breast and uterine tissue of mothers.

Some also warned of children being only partially cured of severe diseases, and posing

formidable, lifelong costs to their families and the healthcare system. And others pointed out that, even in adults, gene therapy has not succeeded.

"I don't know how we can talk about selling it *in utero*," said Mitchell Globus, professor emeritus of obstetrics and paediatrics at the University of California, San Francisco, and the director of scientific and clinical affairs for Applied Imaging Corp. "We don't have a product to sell post-natally yet."

Despite this, many of those present agreed that the impetus to attempt therapy, despite the risks, is compelling for severe diseases that kill before or shortly after birth, and for seriously disabling diseases with no adequate treatments.

"There is a consensus that *in utero* gene therapy will have a definitive role to play in a small set of diseases," said French Anderson, professor of biochemistry and paediatrics at the University of Southern California, and a leading pioneer of human gene therapy.

While some at the two-day conference complained of the lack of animal data, there were also signs of progress with animal experiments. Janet Larson, for instance, a neonatologist and staff scientist at the Alton Ochsner Medical Foundation in New Orleans, Louisiana, described the successful treatment of cystic fibrosis in knockout mice with *in utero* therapy.

Meanwhile, Anderson is preparing human protocols for using *in utero* gene therapy to treat severe combined immunodeficiency and α -thalassaemia. He estimates that they will be ready for submission as soon as three years from now. But he is now modifying the draft proposals after questions were raised by the NIH's Recombinant DNA Advisory Committee last September (see *Nature* 395, 420; 1998).

Activists at the meeting, led by the Council for Responsible Genetics (CRG), argued that if the US government allows the development of *in utero* gene transfer to fight disease, the technique will inevitably end up being exploited for eugenic purposes.

Paul Billings of the CRG, a medical geneticist who is chief medical officer at Heart of Texas Veterans Health Care System in Dallas, blamed "genetic hysteria" for the meeting participants' openness to developing *in utero* techniques.

"The Council for Responsible Genetics questions why we're investing so much time and effort in gene therapy for very rare conditions when prenatal diagnosis, embryo selection and adoption are available," said Billings. He and others also argued that, by providing therapy to younger and younger fetuses, scientists risk playing into an anti-abortion agenda by giving an earlier definition of the beginning of life.

Meredith Wadman

Senator seeks reassurance on NIH budget

[WASHINGTON] A leading US senator is questioning whether the National Institutes of Health (NIH) can effectively spend the \$2 billion, 15 per cent increase it received in the current fiscal year, and whether it should receive a similarly large boost next year.

Senator Pete Domenici (Republican, New Mexico), the chairman of the Senate Budget Committee, says he may hold an unusual committee hearing in the spring to address the issue. "We want to make sure that the large increases are being used effectively and efficiently, not only for the good of science but for the good of the American taxpayer," Domenici said last week.

Last month, at Domenici's instigation, top Budget Committee staff members led by Bill Hoagland, the committee's chief of staff, met Harold Varmus, the NIH director, and other institute directors. Officials from the Congressional Budget Office and the General Accounting Office also attended. Discussions centred on how NIH distributes funds among its institutes, how it sets research priorities, and how it takes public health needs into account in doing so.

"[NIH was] popping up on the radar screen as an outlier" because it received



Domenici: are the NIH's extra billions spent efficiently?

the single largest percentage increase of any agency in 1999, says Hoagland. "We wanted to make sure that the dollars were being used effectively."

Varmus argues that the money is being spent wisely. "I see no difficulty in managing this increase, and subsequent increases of similar size," he says. "It's crucial that we work with a steady growth plan."

He adds that "every institute and centre has a series of very impressive plans" for handling the new money, which is funding "everything from over 1600 new research project grants to the sequencing of the mouse genome."

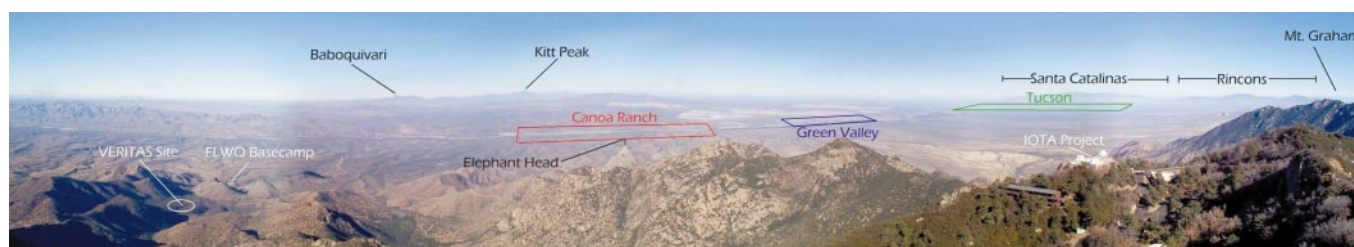
Yesterday (13 January), Hoagland and other Budget Committee staff were scheduled to meet again with Francine Little, the director of NIH's Office of Financial Management. They planned to discuss whether

it is using its \$15.58 billion budget for fiscal year 1999, efficiently, or whether the \$2 billion of new funds is presenting a money management problem.

Hoagland says he wants to know "what plans they have for how they're going to use these monies. Or are they going to be faced with a situation where they cannot manage the resources in a way that is most efficient? In that case I would suggest they are getting too much."

This is strenuously denied by Varmus, who says that "casual" comments that the NIH might be unable to handle the new money are "very damaging to the research enterprise." Varmus adds: "We have no trouble at all in managing the 1999 increase, giving it very careful attention and accounting for every penny." He says he is sure that Hoagland will agree after looking at the evidence.

Separately, unnamed biomedical research lobbyists last week protested at what they say will be a 2.1 per cent, \$328 million increase for the NIH in the White House's proposed budget for the year 2000, due to be released in the first week of February. The research community wants an increase closer to the 15 per cent of 1999. Last year, the Clinton administration asked for 8.4 per cent. **M.W.**



Panoramic view: a 180° view from the summit of Mount Hopkins shows the Canoa Ranch development site 20 miles south of Tucson.

US astronomers contest building project

[WASHINGTON] A dispute over proposed commercial development at the foot of Mount Hopkins in southern Arizona threatens to end years of peaceful coexistence between astronomers and land developers in the state.

Astronomers have opposed the project, fearing that light pollution will degrade viewing conditions at the Whipple Observatory, the Multiple Mirror Telescope (MMT) and other facilities on the mountain.

Fairfield Homes, a developer based in Green Valley, Arizona, wants to expand plans for low-density housing to include commercial development. It has threatened astronomers with a lawsuit if they continue to speak out against the project. Fairfield's application for commercial development on the 5,200-acre Canoa Ranch site 20 miles south of Tucson was scheduled to go before a local board of supervisors at a public hearing this week.

In a letter sent two days before Christmas, Frank Cassidy, an attorney for Fairfield, accused astronomers from the Whipple Observatory and other institutions of lobbying against the project "under the guise of providing scientific information".

Among the "improper actions" cited were the distribution of a brochure on light pollution at an earlier hearing by Whipple public affairs officer Dan Brocius.

Cassidy claimed that, because the Smithsonian Institution observatories are publicly owned, interfering with Fairfield's \$900 million development could amount to a government "taking" of private property, for which opponents of the project would be liable.

Cassidy's letter threatened the institutions as well as individuals — including Robert Kirshner of the Harvard-Smithsonian Center for Astrophysics and MMT observatory director Craig Foltz — with "appropriate legal action" unless they stopped their "lobbying" against the Canoa Ranch application.

Also named in Cassidy's letter was Chris Luginbuhl, an astronomer at the US Naval Observatory's station in Flagstaff, Arizona, and an expert on the impact of lighting on observatories. Luginbuhl was asked by the Whipple Observatory and the University of Arizona (which co-owns the MMT) to provide a critique of Fairfield's estimate of light pollution from Canoa Ranch if commercial development goes ahead.

His estimates were in some cases six to seven times higher than the developer's, based on different assumptions about the types and amount of commercial lots that would be built, and how much light they would produce. In a worst case scenario, says Luginbuhl, commercial development could result in a 10 to 15 per cent brightening of the sky at Mount Hopkins.

Frank Thomson, a planning consultant to Fairfield, says his client is sensitive to astronomers' worries, and is committed to producing no more light pollution than would result from the already approved plan for 1,200 homes.

But Luginbuhl and other astronomers say verbal promises count for little. The plans could change, or Fairfield could sell the property after permission is granted for development to someone with different ideas.

Thomson says astronomers have been unwilling to meet Fairfield representatives to work out strategies for mitigating light pollution. Normally, light pollution can be reduced through measures such as light shielding, or by local ordinances that restrict lighting after certain hours. But astronomers say that may not be possible in this case, given the type and scale of the proposed development. "The developer believes we can hand them a set of lighting ordinances that can make it OK," says Brocius. "If we had a solution, we would have forwarded it."

Luginbuhl agrees that even municipal codes like those in place for Flagstaff, which he says are among the strictest in the world in protecting astronomers' interests, could not preserve viewing conditions on Mount Hopkins if commercial development occurs to the limit allowed by the proposed plan.

The issue will undoubtedly come up later this year when a committee representing both astronomers and developers, co-chaired by Don Davis of the Tucson-based Planetary Science Institute, takes up the matter of revised lighting codes for the Tucson area, which also includes Kitt Peak National Observatory. The codes have been revised several times since being established in 1972.

Thomson says it is "unfortunate" that tensions have escalated over Canoa Ranch after more than 10 years of astronomers and developers working out their differences in a

friendlier way. But astronomers were rankled by what Smithsonian attorney James Wilson calls Cassidy's "inappropriate attempt to intimidate," and what a Tucson newspaper termed "Fairfield's crude threat".

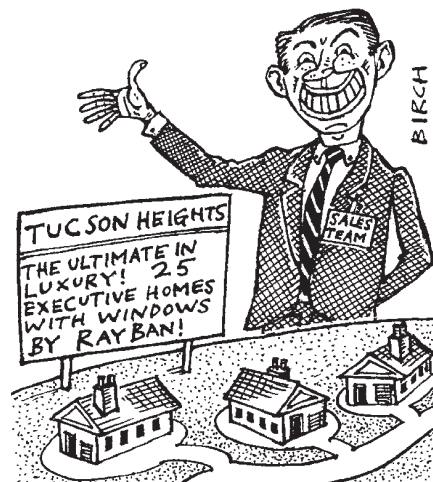
Cassidy's letter advised astronomers to "take your heads out of the sand and recognize the reality that development will inevitably occur in the vicinity of Mount Hopkins".

Wilson says Canoa Ranch is "a big issue for us," and that astronomers on Mount Hopkins have every right to weigh in with their opinions as "affected landowners". As of last weekend, some 1,500 astronomers worldwide had signed a petition opposing "attempts to silence [astronomers] through intimidation". Faced with such criticism, Cassidy was more moderate in a 7 January reply to Wilson, saying "Fairfield has no desire to litigate against the observatory".

Another reason for the acrimony over Canoa Ranch is the high political and financial stakes in the project. Fairfield faced political opposition before this week's hearing, with two of five county supervisors on record as opposing the development application even before the light pollution issue surfaced.

Luginbuhl believes an underlying reason for the heightened conflict is that developers are beginning to encroach on what were until recently considered remote, dark-sky locations. "Astronomers are beginning to wake up to the realization that we can't run any further," he says.

Tony Reichhardt



Fuel additives put under scrutiny — again

[WASHINGTON] The US Environmental Protection Agency (EPA) has asked an expert panel to review possible health and environmental problems associated with the 'oxygenates' that are added to vehicle fuel to help curb air pollution.

The review is the fourth major study of the issue commissioned by the EPA since 1996. It follows a report last November by researchers at the University of California questioning the benefit of fuel oxygenates and calling for the phasing out of the most common additive, a petroleum product known as methyl tertiary-butyl ether (MTBE).

The use of oxygenates such as MTBE and ethanol has been controversial since 1990, when Congress ruled that gasoline should be reformulated to contain at least two per cent oxygen. This was meant to reduce toxic air pollutants and exhaust emissions that contribute to smog. In addition, fuels sold in some areas were required to contain at least 2.7 per cent oxygen during the winter to fight carbon monoxide pollution.

Shortly after MTBE entered wide use in 1992, however, health officials began receiving complaints of headaches, nausea and other adverse effects. That prompted a spate of EPA-requested studies.

A 1996 review of the literature by the Boston-based Health Effects Institute (HEI), which is funded by the EPA and industry, was followed by a National Academy of Sciences study. These fed into an assessment by the White House science office published last year. All three found insufficient evidence to warrant a ban on MTBE, or even a reduction in its use, and called for more research.

In 1997 the State of California, which has its own reformulated gasoline programme, asked researchers at the University of California to conduct another review of the literature on oxygenates. Their report recommends that the state should consider phasing out MTBE over several years and let refiners switch to other types of cleaner-burning gasoline, without the oxygenates.

The conclusion had less to do with new evidence on harmful effects of MTBE — research on the topic remains sparse — than with a judgement that oxygenates contribute little to cleaning up air pollution. Their primary claimed benefit, reducing carbon monoxide emissions, is difficult to prove.

Cleaner-burning cars may be at least as responsible for the decline. In fact, said the University of California researchers, "MTBE and other oxygenates were found to have no significant effect on exhaust emissions from advanced technology vehicles".

They did no better at reducing the toxic air pollutant benzene than did other kinds of reformulated gasoline. But they increased emissions of formaldehyde, known to cause cancer in animals.

In addition, according to the 1996 academy study and others (but not according to the EPA or last year's White House study), oxygenate levels of 2.7 per cent in winter fuels may even add to pollution by producing nitrogen oxides, one of the main offenders in creating urban smog.

The new review will be chaired by HEI president Daniel Greenbaum, who says it will be the first time all interested parties — researchers on air and water quality, health



Smog warning: oxygenates added to winter fuels may increase pollution rather than curb it.

officials, environmentalists and oil industry representatives — will meet to discuss the oxygenates issue. Their report is due in June.

Greenbaum says there have been some new developments since HEI's last report on the subject in 1996, although "we haven't had a lot of brand new scientific information". Additional concerns have been raised about the ingestion of MTBE via drinking water.

The substance percolates easily through groundwater, and some communities have experienced acute episodes of water pollution after underground storage tanks ruptured. The pro-MTBE Oxygenated Fuels Association argues that the leaky tanks, not the oxygenates, are the problem.

Among the many open research questions on MTBE are how it is metabolized, and what the levels of exposure in the general population are. Based on their review, the University of California researchers concluded that MTBE is an animal carcinogen with the potential to cause cancer in humans.

But, according to Greenbaum, three other groups — the International Agency for Research on Cancer, the National Institute of Environmental Health Sciences in North Carolina, and the California Office of Environmental Health Hazard and Assessment — have recently concluded that MTBE is not a likely human carcinogen.

Scientific questions may not ultimately decide the oxygenates issue. Politicians from states with strong farming communities have lobbied for the use of ethanol, a corn-based oxygenate, while representatives from California and Maine have pushed to exempt their states from the federal requirements.

If the additives are phased out, the University of California researchers recommended caution in approving any alternatives. "Addition of any chemical compound to the environment in quantities that constitute a significant fraction of the total content of gasoline may have unexpected environmental consequences."

Tony Reichhardt

Russian physicists may pay price of success

[MOSCOW] Two researchers in the Russian Academy of Sciences have warned that a new tax code being discussed in the Russian parliament "could kill Russian science rather than support it".

The warning has come from Eduard Kruglyakov and Veniamin Sidorov, who work at the Novosibirsk Nuclear Physics Institute (NNPI) of the academy's Siberian branch. The institute is a rare example of a successful and self-financing Russian scientific organization, having lost only 15 per cent of its staff during the hard times of *perestroika*.

Unlike many research institutes, NNPI can still afford equipment and pays salaries regularly, allowing it to continue research in fundamental areas of high-energy physics, the physics of accelerators, plasma and controlled thermonuclear synthesis.

The institute stays in operation primarily through the production and sale

of high-technology equipment, such as particle accelerators, low-dose X-ray machines, and instruments for studying high temperature plasma.

The money collected in this way has made up for the lack of government support. But the new tax code will treat these revenues as profit, and subject to taxation.

"It is hardly possible to equate profit which is spent on buying villas at prestigious resorts, increasing personal bank accounts, or even paying dividends to shareholders, with that which is entirely used for supporting fundamental research," say Kruglyakov and Sidorov.

Writing in the local newspaper of Novosibirsk academic town, which depends entirely on research institutes, they argue that "when the state is unable to finance its science, it has no moral right to tax a scientific organization's profit as if it was just a commercial enterprise".

Carl Levitin

Global 'eco-survey' plan gets a rough ride

[LONDON] A global assessment of the state of the world's ecosystems, in which thousands of scientists worldwide would look at the extent to which ecosystems can continue to support human needs, has been proposed by a group of scientists and environment policymakers.

The proposed assessment is modelled partly on the Intergovernmental Panel on Climate Change (IPCC), an organization of climate scientists set up in 1987 by the United Nations Environmental Programme (UNEP) and the World Meteorological Organization.

But despite its support in the scientific and conservation community, governments of both developed and developing countries have given the proposal a lukewarm reception. There is a feeling that few governments will heed a document whose contents they cannot influence.

The first international scientific assessment of the state of the world's ecosystems would aim to identify 'hot spots' — constituents of the natural environment, such as species, forests and fisheries, that are under threat — and suggest remedial action. It would begin next year and end in 2002.

The overall aim of the so-called Millennium Assessment would be to provide a single source of accurate, policy-relevant ecological science advice to national governments and to UN environment conventions, including those covering climate change, biodiversity, desertification, fish stocks and forests.

The project has received enthusiastic backing — including some promises of financial support — from, among others, the World Resources Institute, the World Conservation Union (IUCN), the World Bank, various United Nations agencies, the International Council for Science (ICSU) and the Megascience Forum of the Organization for Economic Cooperation and Development.

It is also supported by the government of

Ecuador, whose environment minister is president of the IUCN. But representatives of many other governments are less enthusiastic. One specific concern is that conservation groups may try to use the results of such an assessment to influence national conservation policies: at present, these groups are denied a strong voice on the main international conservation decision making body, the UN biodiversity convention.

One representative from a developed country points out that signatories of the UN biodiversity convention have already given their blessing to the more focused — but so far unfunded — Global Taxonomy Initiative, put forward by Diversitas, a network of scientists organized through the United Nations Educational, Scientific and Cultural Organization and ICSU. This aims to do primary taxonomic research, and to train scientists from developing countries.

But, despite unanimous support from governments, funding agencies such as the UN's Global Environment Facility (GEF) appear unlikely to fund the taxonomy initiative. In contrast, the ecosystems assessment has the enthusiastic backing of GEF's chief executive, Mohammed El-Ashry.

Critics of the ecosystems assessment also claim that comparisons with the IPCC are invalid. The IPCC was set up at the request of signatory countries of the UN climate convention for an accurate scientific assessment of the world's climate. Its latest report confirmed a human role in climate change and led to the signing of the Kyoto Protocol (see *Nature* **390**, 649-650; 1997).

But according to John Ashe, Antigua and Barbuda's ambassador to the United Nations, none of the UN environmental conventions has requested an ecosystems assessment. "This looks like scientists dreaming up jobs for themselves," says Ashe. "If such an assessment is needed, it ought to be requested by governments, and not handed down in this top-down way."

Many policymakers from developed and developing countries share Ashe's scepticism. Tewolde Berhan Egziabher, general manager of the Environmental Protection Agency of Ethiopia, says that the IPCC was asked for guidance on human-induced climate change as "there was a major difference of opinion among scientists" on that issue. By contrast, he says, "I don't think that anyone in conservation is saying that species extinction or natural resource depletion is not an important issue."

Both Egziabher and Maurice Iwu, a member of the Nigerian delegation to the biodiversity convention, acknowledge that an ecosystem assessment has scientific merits, "provided that it involves carrying out primary research in the field, and is not a re-hash of what we already know".

Another concern for developing countries is the strong support for the initiative from conservation groups and UNEP. Relations between developing countries and the conservation community, including UNEP, have often been tense. This is partly because the former suspect the latter of promoting conservation at the expense of development.

Six years ago, state parties to the UN biodiversity convention snubbed a UNEP assessment of the state of the world's biodiversity, saying that the agency had failed to obtain their consent or that of their scientific advisory body before undertaking the study.

Abdul Hamid Zakri, the Malaysian chairman of the biodiversity convention's scientific advisory body, says that most developing countries saw the UNEP document as "others telling us what our priorities are". He thinks that the ecosystems assessment will fare better if it is intended as "guidance only and is not in any way an obligatory measure".

But the assessment's backers — particularly conservation groups — are motivated partly by a sense that scientific advisory bodies attached to UN environmental conventions are too political, in that their government-appointed member scientists are swayed unduly by national or regional political priorities when giving scientific advice.

Governments will be invited to nominate the assessment's technical lead authors. But the authors themselves will be chosen on scientific merit by a separate policy committee of representatives of UN and non-government scientific and environmental institutions, with ICSU playing a prominent role.

Social as well as natural scientists will take part in the assessment's design. Walter Reid, director of the Millennium Assessment, promises that if the plan does not generate enough support from policymakers after a one-year pilot, launched this year, the main assessment will be scrapped.

Ehsan Masood



Taking stock: researchers performing a biodiversity inventory in Costa Rica inspect an insect trap.

Five-million-year-old possible hominid unearthed in Ethiopia

[LONDON] An international research team including palaeontologist Tim White of the University of California at Berkeley last week announced the discovery in Ethiopia's Awash Valley of fossils of possibly the earliest member of the human lineage — at around 5 million years in age.

According to a report in the *Ethiopian Herald*, White described the find as "extraordinary" during a meeting at Addis Ababa University. White uncovered the 4.4-million-year-old *Ardipithecus ramidus* in the same area (see *Nature* 371, 306–312; 1994).

A full assessment of the importance of the discovery will depend on the age and completeness of the fossil; a skull alone may not give enough information to determine whether the ancestor was bipedal, and thus suggest it is a hominid. The fossil skull of a hominid believed to be 2.5 to 3 million years old has also been found at the site.

Césarsky takes on top job at Euro observatory

[PARIS] The European Southern Observatory in Garching, near Munich, has confirmed the

appointment of French astrophysicist Catherine Césarsky as its new director-general, replacing Riccardo Giacconi (see *Nature* 396, 716; 1998). She is currently director of the department of 'science of matter', a fundamental research division of more than 3,000 researchers and engineers, at the French Atomic Energy Commission.

Césarsky's own research is on modelling of stars and interstellar space. But she has herself recently had a close encounter with models and stars: a full-page portrait of her appeared in last month's French edition of *Vogue*, which was dedicated to the theme "science and fashion".

Monsanto concession on engineered corn

[WASHINGTON] The Monsanto company announced last week that growers of genetically engineered *Bt* corn will have to plant accompanying plots of non-engineered corn that are 20 per cent of the size of the engineered crop. The announcement was made on behalf of a coalition of producers of *Bt* corn, of which Monsanto is the largest.

The companies have been under pressure to introduce such a measure, which in theory reduces pest resistance to *Bt*, the insecticide produced by the engineered corn. The US Environmental Protection Agency has

welcomed the move in principle, but environmentalists have criticized the 20 per cent requirement as insufficient, saying that larger non-engineered plots are needed.

Molecular medicine trio share Swiss prize

[GENEVA] The 1999 Louis-Jeantet Prize for medicine has been awarded to three researchers working in molecular medicine. Adrian Bird, from the University of Edinburgh, and Herbert Jäckle, from the Max-Planck Institute of Biophysical Chemistry in Göttingen, win the award for their work on gene expression.

The third recipient is Jean-Louis Mandel, from the University of Strasbourg Medical School, who works on the diagnosis of genetic lesions causing inherited diseases. This year's winners will each receive SFr600,000 (US\$430,000) to pursue research and personal awards of SFr100,000.

Japan creates mouse DNA encyclopaedia

[TOKYO] The Genome Research Centre, part of Japan's Institute of Physical and Chemical Research, announced on Monday (11 January) that its genome exploration group has created an 'encyclopaedia' of more than

20,000 partial sequences of full-length mouse complementary DNA clones.

The sequence data, from 107,088 cDNA clones, include approximately 5,000 unknown partial sequences identified by comparing sequences with those registered in various human and mouse databases.

Full-length cDNA clones have been sequenced using an automated high-speed sequencer developed by the research group. The 20,000 partial sequences in the Mouse Encyclopedia Index, which accounts for 20 per cent of the entire genome, will be posted on the web at: <http://genome.rtc.riken.go.jp>.

Brussels research boss pledges to stand firm

[BRUSSELS] **Edith Cresson, the European commissioner responsible for education and research, has vowed to resist pressure from the European Parliament to step down in the wake of allegations of poor administrative practices involving Leonardo, a vocational training programme.**

The demands for Cresson's resignation follow a critical report of the programme by the commission's financial control unit. Similar demands are being made of other commissioners over separate allegations of financial mismanagement of the commission's 92 billion euros (US\$107

billion) annual budget. However Cresson, who was appointed to her post in 1994, says that she refuses to be a "human sacrifice" and to give in to "intimidation tactics".

Patent office cuts fees to encourage inventors

[MUNICH] The European Patent Office in Munich is to reduce its registration fees by almost 20 per cent, to DM7,400 (US\$4,400). According to its president, Ingo Kober, the move — the second reduction in fees since 1997 — is intended to encourage the patent activities of individual inventors, research institutes and small companies.

But patents will still remain expensive compared with the US or Japanese patent offices. This is because registration fees account for only around 15 per cent of the application costs, which average DM60,000, and include fees for translation, attorneys and the annual renewal fees that applicants must pay to their national patent offices.

NIH streamlines grant applications process

[WASHINGTON] The US National Institutes of Health (NIH) will begin on 1 June to require research grant applicants to estimate their costs in \$25,000 increments, in so-called

'modular' grant applications (see *Nature* 394, 306; 1998). It hopes that the change will simplify and streamline the grants process by decreasing both the time required of investigators in drawing up budgets and the time taken by peer-review committees in assessing those budgets.

The modular form will apply to all applications for research project grants (R01s), as well as to small (R03) and exploratory/developmental (R21) grants.

Allergy scientists win King Faisal award

[LONDON] Patrick Holt, of Australia's National Health and Medical Research Council, and Stephen Holgate, from Britain's University of Southampton, have been jointly awarded the King Faisal International Prize for Medicine for their work on respiratory allergies.

The chemistry prize was shared by Ryoji Noyori, from Nagoya University in Japan, and Dieter Seebach from the Swiss Federal Institute of Technology. The researchers were honoured for developing methods for the preparation of organic chiral molecules and for achieving selective and efficient chemical synthesis. The joint winners of each prize receive a gold medal and share of US\$200,000 prize money.

French reforms would disrupt research

Sir—The French scientific community is seriously worried about a furtively proposed government decree on the Centre National de la Recherche Scientifique (CNRS). Claude Allègre, the science minister, has proposed a sweeping reform without consulting the organization's director general, the scientific directors, or the members of its scientific committees.

Even in the best scenario, Allègre's proposal will lead to the break up of the CNRS, effectively ending its involvement in formulating research policy (see *Nature* **396**, 607; 1998). The CNRS would become a funding agency ruled by the short term political aims of the government.

As disturbing is the apparently seductive notion that the break up of the CNRS will bolster French universities by entrusting them with the principal responsibility for scientific research. In this utopian plan, a relatively small number of CNRS researchers would be dispersed among the much larger university population with the aim of 'improving' these institutions. A

much more likely outcome will be the dissipation of the research community without significantly affecting the universities. Allègre, fascinated by the US research establishment, seems to forget that French universities, faced with the hard reality of mass teaching, are not elite institutions like the Massachusetts Institute of Technology or Caltech. He also tends to forget that public organizations exist in the United States (the National Institutes of Health), the United Kingdom (Medical Research Council) and Germany (Max Planck Institutes), and that these institutions have an important role in establishing science policy.

Belgium and Italy have experimented with proposals similar to those of Allègre, and we all know how catastrophic this has been. Interference by the state in planning and recruitment has nearly destroyed the research community.

Never before have the CNRS and its researchers been so involved in working with the French universities. As an

independent research organization, the CNRS has had a pivotal role in developing regional research centres affiliated to universities. The CNRS would not have been able to do so if it had been integrated into the university system. So it is all the more strange that a project breaking up the CNRS has been proposed now.

Other research organizations such as the biomedical agency INSERM and ORSTOM, the research agency for developing countries, are likely to be next on the list to suffer from the same measures.

Our aim is to convince Allègre that the proposal needs serious reconsideration. This poorly thought out reform is not the solution, because it will inevitably disrupt the research community. French scientists have an important role in reforming the system, and they deserve to be heard.

Alain Bucheton
(President)

Section 23 — Génomes, fonctions et régulations, Comité National de la Recherche Scientifique, CNRS, 3 Rue Michel-Ange, 75794 Paris Cedex 16, France

Lab guidelines aim to prevent misconduct

Sir—In response to your editorial, "Surviving misconduct is one thing, accountability is another"¹, and a subsequent comment², I would like to bring to your notice an initiative already under way to provide "good guidelines for the conduct of research".

In 1993 the Danish Committee on Scientific Dishonesty, appointed by the Danish Medical Research Council, defined guidelines for the presentation of experimental reports, and data documentation and storage. It was felt that, if such proposals were to have a chance of being followed, they had to be concise and readily available in the research milieu. After consultation with all major medical research institutions in Denmark, the revised guidelines were printed on one sheet of paper — those for basic health research on one side, and those for clinical research on the other side. The page was embedded in acid-proof plastic to prevent it from being damaged in the laboratory.

About 1,000 copies were distributed to laboratories all over Denmark. After a year's trial in 1995, new comments were invited from the same institutions. The guidelines had been well received and were being used. On the basis of proposals for improvement, an updated version was distributed to PhD

students and their supervisors, and made available to all interested parties. This revised version was published in English in a paper describing the motives for the choice of guidelines and how they were adapted to medical research in Denmark³. The guidelines are also available on the web (<http://www.forskraad.dk/publ/guide.html>).

I think this preventive initiative is much in keeping with the proposed "adoption of a standardized recording protocol"².

Torben Clausen

Department of Physiology, University of Aarhus, DK-8000 Århus C., Denmark

1. *Nature* **395**, 727 (1998).

2. Angelides, K. & Pianelli, J. V. *Nature* **396**, 404 (1998).

3. Clausen, T. & Riis, P. *Dan. Med. Bull.* **44**, 85–87 (1997).

Did agriculture cause the population explosion?

Sir—I was pleased to see Thomas Malthus commemorated in Roger Short's bicentennial essay¹, but disappointed that agriculture's role in the population explosion was ignored.

The two "major factors" identified by Short (erosion of traditional breastfeeding practices, and conquest of infectious diseases) did surely contribute to the rapid growth of population, but only because they were accompanied by an increase in the food supply. The expansion of agricultural land since Malthus's time, and

the enormous increases in crop yields resulting from the application of science to agricultural technology, were what allowed the population to grow to 6 billion. Yields per hectare of several major crops have increased by factors of two to seven in the United States in the past 60 years², and in other countries by similarly large factors³.

Malthus's emphasis on food as the limit to population may be confirmed if agricultural productivity soon begins to level off, as many predict³.

Stephen G. Warren

Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195-1640, USA

Short replies — I do not agree that agricultural improvements have had a major impact on human fertility. Women have to be severely malnourished before fertility is affected⁴. State of nutrition is not normally one of the principal determinants of human fertility; the women in Belsen and in the Dutch famine of the Second World War continued to get pregnant.

However, improvements in agricultural productivity have certainly prevented deaths from starvation in today's overpopulated world, and, as Malthus said, at the end of the day food is necessary to the life of the man.

Roger Short

Royal Women's Hospital, 132 Grattan Street, Carlton, Victoria 3053, Australia

1. Short, R. *Nature* **395**, 456 (1998).

2. Warren, G. F. *Weed Technology* **12** (in the press).

3. Mann, C. *Science* **277**, 1038–1043 (1997).

4. Bongaarts, J. *Science* **208**, 564 (1980).

Traffic lights at the Beringian crossroads

Andrei Sher

The first opening of the Bering Strait would have had profound biogeographical and climatic consequences. The date of that event is now firmly pushed further back in time.

Beringia is the area where the two greatest landmasses on Earth (Eurasia and America) and two great oceans (the Pacific and Arctic) meet. At the moment, the two continents are separated by the Bering Strait, providing a green light for interchange of marine dwellers but preventing most terrestrial organisms from travelling between the Old and New Worlds. The lights, however, changed only about 11,000 years ago, when world sea level recovered from the last deep marine regression caused by global cooling and the locking-up of water in continental glaciers.

Before that, the opening of land or marine migration routes switched repeatedly, ever since the tectonic separation of the two landmasses which, before the Neogene (Fig. 1), had constituted a single continent. The exact time when this controlled junction started to operate is still not clear; but evidently, by then, the polar regions were already much colder than the equatorial ones, and the opening of a new route of heat transfer between high and low latitudes would have been a momentous event. So, besides making migration of marine animals possible, and that of terrestrial animals much less likely, the first connection between Pacific and Arctic waters must have had an enormous impact on climate in the Northern Hemisphere.

The age estimates of this occurrence vary from 3–4 million years ago (Myr) to 8–10 Myr or even older. Clues for more precise dating lie in the Bering and Chukchi Seas and the Bering Strait. This region, which separates two great and not always friendly powers, has been a traditional area of cooperation, and one such collaborative endeavour is described on page 149 of this issue¹. There, Marinovich and Gladenkov report new evidence of the date by which the Bering Strait was certainly open. The evidence comes from the Alaska Peninsula, and is based on a study of marine fossils from the 276-metre-thick Sandy Ridge section, complemented by earlier work from Karaginsky Island on the Russian side of the Bering Sea (see map on page 150). The clear signal for the Bering Strait's opening is the appearance of the

Atlantic–Arctic bivalve mollusc *Astarte* in sediments of the North Pacific Basin, independently dated from microfossils of diatoms, marine algae, found in the same sediment layers.

The diatoms of well-established age² allow Marinovich and Gladenkov to determine the minimum age of *Astarte* immigration from the Arctic as 4.8–5.5 Myr. However, they report that *Astarte* occurs much further down the Sandy Ridge sequence, implying that it may have appeared in the Pacific even before then. On the Russian side, its first occurrence is in the Karaginsky Island section, where it appears within a formation that presumably has a long timespan but unfortunately lacks a diatom record. The nearest level where diatoms are found (together with *Astarte*) up the Karaginsky section can be dated as 3.4–4.1 to 4.8 Myr; the nearest diatom assemblage down the section (without *Astarte*) could be as old as 7.3–7.4 Myr. Marinovich and Gladenkov take the latter date for the maximum possible age for the first opening of the Bering Strait, and this is probably the only weak point in an otherwise exciting story.

One immediate conclusion is that Arctic marine mollusc faunas containing Pacific species can no longer be considered as younger than 3–4 Myr, as it is commonly believed now, which has practical implications in defining the age of former high sea-level stands on Arctic coasts. So Marinovich and Gladenkov have established a landmark at the Beringian crossroads, indicating that the route for marine migration from the Arctic to the Pacific certainly existed as early as 4.8–5.5 Myr (that is,

around the Miocene–Pliocene transition).

But the first opening of this route could be earlier. Indeed it could be much earlier, if we turn to evidence from the northern segment of the route, where the first appearance of Pacific organisms in Arctic waters can indicate the strait's opening. On the Arctic coast of the Chukotka Peninsula, sediments contain various, predominantly marine, diatoms³. Some of them are age-diagnostic species of the North Pacific in the Late Miocene, and they occur in a narrow time-range between 10 and 9 Myr (ref. 2). Further west, other sediments contain even older Pacific diatoms, known from the end of the Middle to the beginning of the Late Miocene⁴. Based on her study of diatoms from North Chukotka, Polyakova⁴ postulates that the Arctic–Pacific marine connection appeared as early as around the transition between the Early and Middle Miocene (that is, at 16–17 Myr), and has periodically existed since the end of the Middle Miocene (11–12 Myr). But the Chukotka record, lacking molluscs, is still open to debate, and in this respect the evidence from Sandy Ridge looks more solid.

What would have been the consequences of the first opening of the Bering Strait—the red light at the bridge—for the terrestrial life of Eurasia and America? That is no easy question. There is a tradition of estimating periods of faunal exchange by the times of the first appearance of immigrants from the other side, usually based on comparison of mammalian records in the lower latitudes of the two continents. But such correlations, useful as they are, can only be provisional because they are not corroborated by records from the huge area across which the migration route ran: the high-latitude record is vanishingly poor.

During the Pliocene, for example, many important mammalian migrations are postulated, and they can be provisionally pigeonholed into periods of low sea level in the global scale⁵, or into the supposed breaks between three or more marine transgressions in Beringia⁶. But the general climatic cooling in the Pliocene was particularly advanced in high latitudes, so it remains unclear how the rather rigorous environmental conditions in the broad Beringian area could have allowed many of these dispersals.

Meanwhile, when new palaeontological evidence about early mammals appears in

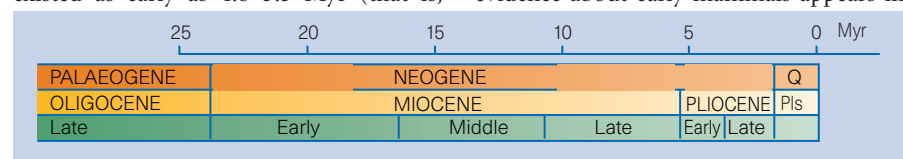


Figure 1 Timescale for events in Beringia, in millions of years (Myr). Before the Neogene, Eurasia and America constituted a single continent, and there was no marine connection between the Pacific and Arctic Oceans. Since some point in the Miocene, periodic rises and falls in sea level have resulted in the opening and closing of the Bering Strait. In the work discussed here, Marinovich and Gladenkov¹ date the first opening to at least 4.8–5.5 Myr. Q, Quaternary Period; Pls, Pleistocene. (Modified from ref. 5.)

Beringia and adjacent regions, it is usually not quite what we would expect from existing dispersal models. So far, discoveries of Late Cenozoic Beringian faunas and floras indicate that the climate was rather cool and highly continental, and they show pronounced Eurasian rather than North American affinities, even if found east of the strait. This is exemplified by the discovery of hipparion (three-toed-horse) fossils on Ellesmere Island in the Canadian Arctic⁷. In the Neogene, these animals were among the most numerous and diverse hoofed mammals in both the New and the Old Worlds, but they had never previously been found in Beringia, or indeed anywhere at high latitudes. The Ellesmere Island fossils, which supposedly date from the Early Pliocene (3–4 Myr), do not look like any North American species; instead they resemble some Chinese forms, in agreement with the biogeographical implications of accompanying fossils. Similarly strong Eurasian connections are evident in the later Pliocene and Early Pleistocene fossil mammal assemblages from east Beringia⁸.

The more we find out about the terrestrial life of Beringia in the Late Cenozoic, the less justified is the view of the land bridge as an intercontinental highway with traffic lights. It gives place to the concept of a Beringian Land (or Beringida) as an evolutionary centre, where many characteristic Pleistocene species and communities evolved and became adapted to generally cold and dry climate with sharp seasonal contrasts and permafrost. Most of Beringida can be seen as an extension of the vast and shallow East Siberian shelf, and that, together with its peculiar climate, may explain why its main evolutionary lineages are rooted in Asian stocks.

Many continental elements of the Beringian biota could have hardly existed close to the cold Arctic sea waters. Various lines of evidence indicate that for later Pleistocene times a relatively narrow zone of the Bering isthmus, with its wetter environment, could have formed a barrier to the dispersal of many animals adapted to dry conditions, especially among the terrestrial invertebrates, even during moderate regression stages^{9–12}.

The importance of the first opening of the Bering Strait lies not only in the splitting of Beringida in two, and the consequent opening and closing of migration routes, but also in the effect it would have had on the environment of this large, high-latitude subcontinent. The establishment of water flow between the Pacific and Arctic Oceans would probably have strongly influenced general ocean circulation, and hence global climate (a topic now being tackled by modelling); but it must have had a bigger impact on the regional climate and so on the terrestrial life that had existed there for millions of years. The open connection between two huge

water masses with different temperatures would, in all likelihood, have resulted in more turbulent atmospheric conditions, increased precipitation and lower summer temperatures. Combined with the generally low temperature at these high latitudes, the resulting increased humidity could eventually have led to the emergence of treeless tundra-like communities and their wide dispersal over adjacent areas.

During the past 30 years of Beringian studies, the age of many remarkable events in the region has almost always been revised down the timescale. That is a striking trend, emphasizing the great antiquity of the Beringian realm. The results reported by Marinovich and Gladenkov are a further and notable step in that direction, but far from the end of the story. □

Andrei Sher is in the Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninskiy Prospekt, Moscow 117071, Russia.

e-mail: asher@glas.apc.org

1. Marinovich, L. Jr & Gladenkov, A. Yu. *Nature* **397**, 149–151 (1999).
2. Barron, J. A. & Gladenkov, A. Yu. *Proc. ODP Sci. Res.* **145**, 3–19 (1995).
3. Zyryanov, E. V., Laukhin, S. A. & Polyakova, E. I. *Izvest. AN SSSR, Ser. Geol.* **2**, 48–55 (1992) (in Russian).
4. Polyakova, E. I. *The Eurasian Arctic Seas During the Late Cenozoic* (Nauchny Mir, Moscow, 1997) (in Russian).
5. Woodburne, M. O. & Swisher, C. C. III *Soc. Sediment. Geol. Spec. Publ.* **54**, 336–364 (1995).
6. Basilyan, A. E., Barinov, K. B., Oreshkina, T. V. & Trubikhin, V. M. in *Paleogeography and Biogeography of the Pliocene and Anthropocene*, XIII INQUA Congress 5–24 (Geol. Inst. AN SSSR, Moscow, 1991) (in Russian).
7. Hulbert, R. C. Jr & Harington, C. R. *J. Vert. Paleont.* **16** Suppl., 42A (1996).
8. Repenning, C. A. & Browsers, E. M. *US Geol. Surv. Bull.* **2036**, 37 (1992).
9. Berman, D. I. & Alfimov, A. V. in *Beringian Paleoenvironments Workshop* 32–33 (Florissant, Colorado, Prog. Abstr., 1997).
10. Elias, S. A. in *Beringian Paleoenvironments Workshop* 56–59 (Florissant, Colorado, Prog. Abstr., 1997).
11. Guthrie, R. D. in *Beringian Paleoenvironments Workshop* 67 (Florissant, Colorado, Prog. Abstr., 1997).
12. Elias, S. A., Short, S. K., Nelson, C. H. & Birks, H. H. *Nature* **382**, 60–63 (1996).

Cell cycle

Checkpoint on the nuclear frontier

Jonathon Pines

Space is set to be the next frontier in cell-cycle research. Until now, the focus has been on how the cell cycle is regulated in time, by the synthesis and destruction of regulatory proteins such as the cyclins. But this is a one-dimensional view, and there is increasing evidence that the cell cycle is controlled by localizing specific regulators to the right place at the right time. An example of this is reported by Paul Russell and colleagues¹ on page 172 of this issue. They show that, in the fission yeast *Schizosaccharomyces*

pombe, the Cdc25 phosphatase — an essential regulator of mitosis — is exported from the nucleus in response to DNA damage. Export separates Cdc25 from its substrate, the cyclin B/Cdc2 kinase, thereby stopping cells from entering mitosis. Moreover, Cdc25 is exported from the nucleus by association with a protein called Rad24, which acts as an attachable nuclear-export sequence.

To maintain the integrity of the genome, DNA must be fully replicated and undamaged before mitosis can begin. So, the initia-

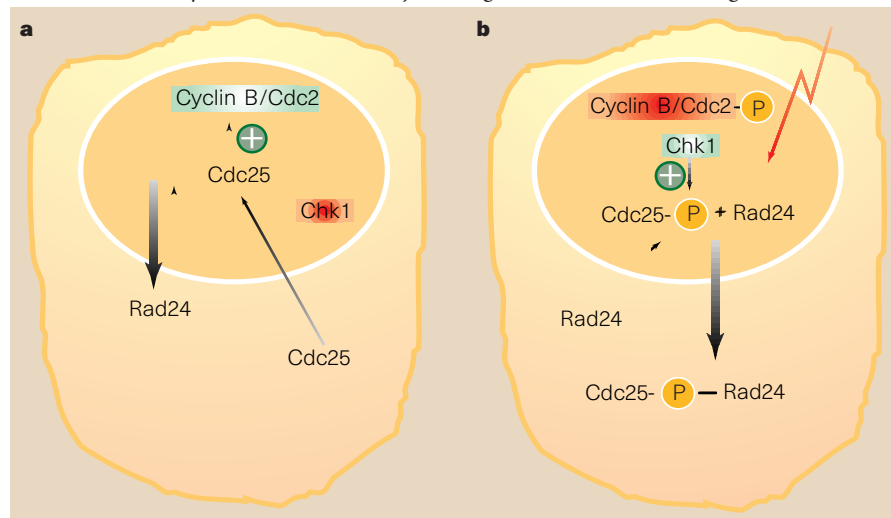


Figure 1 Regulation of mitosis in the fission yeast *Schizosaccharomyces pombe*. a, Normally, Cdc25 accumulates in the nucleus and activates cyclin B/Cdc2 to trigger mitosis. Chk1 is inactive, and Rad24 shuttles in and out of the nucleus, although its export is faster than its import. b, In irradiated cells, DNA damage activates Chk1, which then phosphorylates Cdc25. Russell and colleagues¹ have found that this creates a phosphoserine binding site for Rad24, which exports Cdc25 out of the nucleus. Cyclin B/Cdc2 remains in its phosphorylated, inactive state and the cell cannot enter mitosis until the damaged DNA has been repaired.

tion of mitosis is carefully regulated, and the essential elements have been conserved through evolution². The main target of this regulation is cyclin B/Cdc2. During interphase, this complex is kept inactive by the wee1 kinase, which phosphorylates Cdc2 on a conserved tyrosine residue (Y15) in the ATP-binding site. To initiate mitosis, the phosphate is then removed by Cdc25. DNA damage prevents cells from entering mitosis by enhancing phosphorylation of Y15, so keeping the cyclin B/Cdc2 turned off. This arrests the cells in the G2 phase at the 'DNA-damage checkpoint' (reviewed in ref. 3).

If the DNA-damage checkpoint is compromised, cells become very sensitive to radiation because they begin mitosis with damaged DNA. This sensitivity was used as the basis of a genetic screen to isolate components of the checkpoint in fission yeast. Among other proteins, the screen identified the Chk1 serine/threonine kinase and Rad24, which is a 14-3-3 protein — a family of proteins that recognize and bind to phosphoserine in specific contexts⁴.

Further genetic studies indicated that Chk1 and Rad24 act through Cdc25. In both human cells and fission yeast, damaged DNA activates Chk1, which then phosphorylates Cdc25 on a conserved serine residue (S216), creating a phosphoserine binding site for a 14-3-3 protein⁵⁻⁷. A mutant Cdc25 with an alanine residue (which cannot be phosphorylated) at position 216 cannot bind a 14-3-3 protein, and cells expressing this mutant have a defective DNA-damage checkpoint⁶. However, Cdc25 seems to have the same activity *in vitro*, whether or not it is bound by a 14-3-3 protein. How, then, does Rad24 inactivate Cdc25 in the cell?

The first clue came from more studies on human cells. The 14-3-3 σ protein, which is strongly induced by radiation damage just before mitosis, seems to exist exclusively in the cytoplasm⁸. In contrast, human Cdc25 is found mainly in the nucleus. Thus it was suggested that the 14-3-3 proteins might prevent mitosis by sequestering Cdc25 in the cytoplasm⁹.

This model has now been validated, at least in fission yeast, by the studies of Russell and colleagues¹. They show that Cdc25 is mostly cytoplasmic whereas its substrate, cyclin B/Cdc2, is always nuclear. As the cell progresses towards mitosis, some Cdc25 accumulates in the nucleus, suggesting that mitosis is triggered when the activity of the nuclear Cdc25 exceeds that of wee1. When the authors irradiated fission yeast, Cdc25 remained in the cytoplasm and the cells did not enter mitosis. But in cells that lacked either Chk1 or Rad24, Cdc25 continued to accumulate in the nucleus after irradiation, and the cells entered mitosis with damaged DNA. Chk1 is a nuclear protein, suggesting that it normally phosphorylates the nuclear pool of Cdc25, creating a binding site for

Rad24. As in human cells, however, Rad24 seems to be cytoplasmic. But with a relative molecular mass of 30,000, Rad24 is theoretically small enough to diffuse through the nuclear pore. So is Rad24 actively excluded from the nucleus?

To test this, Russell and colleagues used a strain of fission yeast defective in Crm1/exportin1, which exports proteins from the nucleus. They found that, in these cells, Rad24 was not excluded from the nucleus. The cells were also sensitive to irradiation. Moreover, Cdc25 remained in the nucleus after irradiation, pointing to a connection between the nuclear export of Rad24 and that of Cdc25. Proteins that are exported by Crm1 usually have a hydrophobic region, which acts as a nuclear-export signal (NES). Cdc25 does not have a recognizable NES, but Rad24 does. Gratifyingly, when the authors mutated the Rad24 NES, both Rad24 and Cdc25 were not exported from the nucleus, and the DNA-damage checkpoint was lost. So, Rad24 seems to act as a phosphorylation-dependent attachable NES for Cdc25, regulating mitosis by separating Cdc25 from cyclin B/Cdc2 (Fig. 1).

Russell and colleagues' study not only shows the contribution of the spatial dimension in controlling the cell cycle, but also underlines the importance of regulated nuclear import and export. In human cells, cyclin B/CDC2 shuttles between the nucleus and the cytoplasm. Then, at the end of prophase, it translocates into the nucleus owing to a phosphorylation-dependent shift in the balance between its import and export¹⁰⁻¹². Similarly, human CDC6 and

cyclin D, which help to regulate DNA replication, are exported into the cytoplasm after phosphorylation^{13,14}. For cyclin D this also correlates with destruction by ubiquitin-dependent proteolysis¹⁴, and the connection between nuclear export and proteolysis has also been made¹⁵ for p53. Moreover, in budding yeast, some mutants that are defective in nuclear transport cannot degrade mitotic cyclins¹⁶. Ubiquitin-dependent proteolysis is fundamental to cell-cycle control, so the emerging connection between ubiquitination and nuclear transport may be important for cell division. Clearly, however, for a complete understanding we must study cells in four dimensions. □

Jonathon Pines is at the Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK.
e-mail: j.pines@welc.cam.ac.uk

1. Lopez-Girona, A., Furnari, B., Mondesert, O. & Russell, P. *Nature* **397**, 172–175 (1999).
2. Nurse, P. *Nature* **344**, 503–508 (1990).
3. Elledge, S. J. *Science* **274**, 1664–1672 (1996).
4. Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. *Cell* **84**, 889–897 (1996).
5. Furnari, B., Rhind, N. & Russell, P. *Science* **277**, 1495–1497 (1997).
6. Peng, C. Y. et al. *Science* **277**, 1501–1505 (1997).
7. Sanchez, Y. et al. *Science* **277**, 1497–1501 (1997).
8. Hermeking, H. et al. *Mol. Cell* **1**, 3–11 (1997).
9. Nurse, P. *Cell* **91**, 865–867 (1997).
10. Hagting, A., Karlsson, C., Clute, P., Jackman, M. & Pines, J. *EMBO J.* **17**, 4127–4138 (1998).
11. Toyoshima, F., Moriguchi, T., Wada, A., Fukuda, M. & Nishida, E. *EMBO J.* **17**, 2728–2735 (1998).
12. Yang, J. et al. *Genes Dev.* **12**, 2131–2143 (1998).
13. Petersen, B. O., Lukas, J., Sorensen, C. S., Bartek, J. & Helin, K. *EMBO J.* (in the press).
14. Diehl, J. A., Zindy, F. & Sherr, C. J. *Genes Dev.* **11**, 957–972 (1997).
15. Roth, J., Dobbela, M., Freedman, D. A., Shenk, T. & Levine, A. *EMBO J.* **17**, 554–564 (1998).
16. Loeb, J. D. et al. *Proc. Natl Acad. Sci. USA* **92**, 7647–7651 (1995).

Atomic physics

A new twist in β -decay

Wolfgang Stoeffl

In an elegant experiment, described on page 137 of this issue, Gatti *et al.*¹ have found a new twist in the β -decay of a rhenium isotope. Radioactive β -decay is when a neutron in an atomic nucleus changes into a proton, emitting an electron and a neutrino in the process. The authors show that the local molecular environment of the decaying nucleus has a significant effect on the shape of the electron emission spectrum. The underlying interest in β -decay, and the rationale for Gatti and colleagues' experiment, stems from the hunt for that elusive particle, the neutrino.

Some 70 years ago, the β -decay emission spectrum was a real puzzle. The emitted electrons carry away an energy somewhere between zero and the full available decay energy, unlike the characteristic decay energies of other radioactive particles. The missing energy in β -decay was such a mystery that for a while even Einstein, like many oth-

ers, accepted the idea that energy was not conserved in this process. To solve the problem, Pauli introduced the neutrino in his famous 1930 letter to the "Dear Radioactive Ladies and Gentleman". He also suggested that, "like the new taxes", it's best not to think about it². But it did not take long before the idea of the neutrino was accepted.

When Fermi formulated the theory of β -decay in 1933 (ref. 3), he already recognized that the shape of the emission spectrum could be used to study the mass of the neutrino. Even then, it was known that the neutrino is a very elusive, light particle. A neutrino mass changes the shape of the β -spectrum near the high-energy endpoint. If the neutrino has mass, the electron cannot carry away the full decay energy, and the electron spectrum bends downward just before the endpoint. The detailed shape of this tiny cutoff is a measure of a possible neutrino mass.

Over the years, the theory of neutrinos

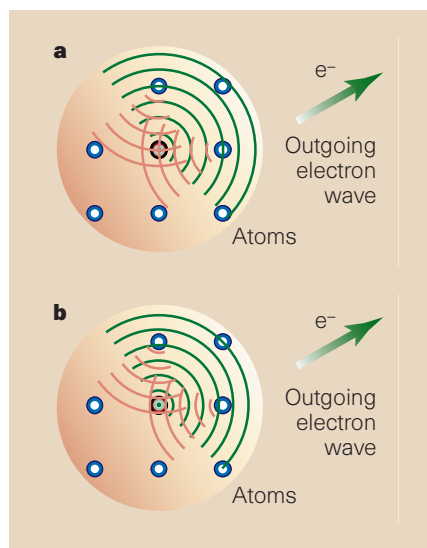


Figure 1 The interaction between the emitted electron in β -decay and its local atomic environment. The explanation for 'β-environmental fine structure' (BEFS) is the wave structure of the electron. In a, an outgoing electron wave of a certain energy is backscattered off neighbouring atoms and constructively interferes with itself at the decaying nucleus; in b, a wave of slightly higher energy (or shorter wavelength) destructively interferes. The various distances to the neighbours create a complex spectrum for different electron wavelengths or energies.

was refined and there were plausible arguments that they should have zero mass. The Standard Model of modern physics, which describes nearly all of the observed interactions between elementary particles, requires a massless neutrino, but there are a few extensions to the model that allow it to have a tiny mass. Big Bang theory tells us there should be about a thousand million times more neutrinos than protons and neutrons in the Universe. So, even if the elusive neutrino was minuscule, taken together neutrinos would make the Universe a lot heavier than all the stars combined. The neutrino is a candidate for some of the missing 'dark matter' between the stars and galaxies, material we cannot see except by the pull of gravity.

Here we come back to Gatti and colleagues' experiment¹. To measure the mass of the neutrino, one has to study the energy spectrum of the electron emitted in β -decay with great precision. There is no known way to measure its mass directly; it can pass through 100 light years of lead without interaction. Many different types of detectors have been developed to measure the energy of the decay electron in the past 20 years. Some groups used electric or magnetic fields, and others have developed new kinds of cryogenic detectors.

At temperatures close to absolute zero, the specific heat of some crystal materials drops to such a low value that the tiny heat pulse

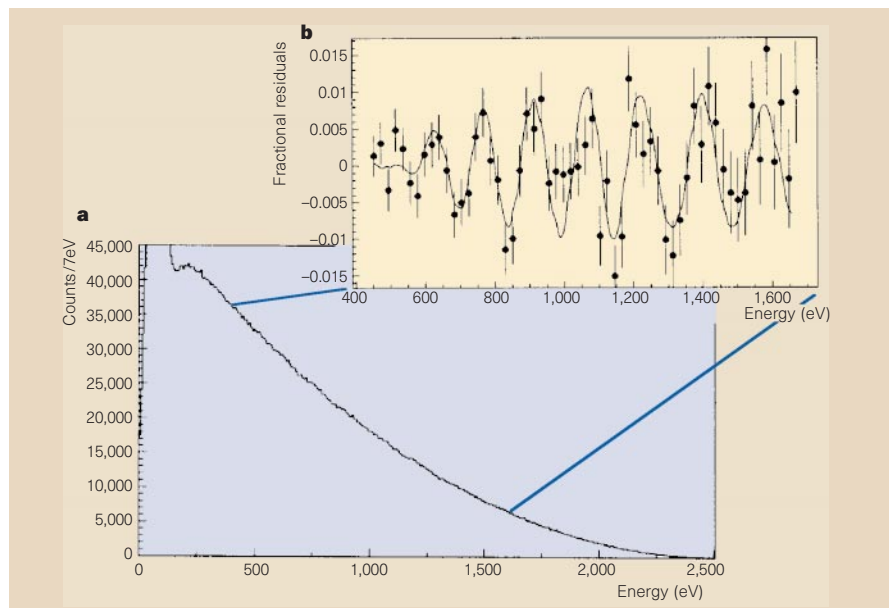


Figure 2 The overall shape and fine detail of the ^{187}Re β -decay spectrum measured by Gatti *et al.*¹. a, The complete β -spectrum of ^{187}Re . b, β -environmental fine structure, which is expressed as a fine ripple on top of the standard β -spectrum. The smooth line shows the predicted shape of the BEFS spectrum. At low energies, the ripples are too close together to be resolved by the detector, and at high energies the oscillations become negligible.

released from a single electron impact increases the temperature of the millimetre-sized crystal by a measurable amount. With a fast thermometer, one can precisely measure the energy released by a single radioactive decay. In the past few years, the new cryogenic detectors have become the best tools available to study low-energy β -decay spectra with high resolution. To get the best sensitivity for detecting a possible neutrino mass, one looks for isotopes with the lowest possible decay energy. For this reason Gatti and coworkers chose to study the decay of rhenium (^{187}Re).

Before one can identify a limit for the mass of the electron-neutrino (none has been reliably observed so far), one has to understand all the little perturbations to the shape of the spectrum caused by the environment of the nucleus. For example, the simple bell-shaped curve of β -decay has to be corrected for the Coulomb force, which makes the spectrum very asymmetric towards low energies. This has been known for many years, but it was not until 1991 that Koonin⁴ recognized that there should be structure in the spectrum depending on the atomic and molecular surrounding of the decaying nucleus. He called it 'β-environmental fine structure' or BEFS.

This effect is analogous to the well-known 'extended X-ray absorption fine structure' used to study molecular structure in solids⁵. There is little difference between how the outgoing electron is created, whether by the absorption of an X-ray in an atom or by nuclear decay. When the electron tries to leave the atom, it can be reflected from the neighbouring atoms to produce a 'standing wave' at the original atom (Fig. 1). Depending on the distance to the next atom-

ic neighbours and the energy of the electron, the wave structure of the electron can increase or reduce the available 'phase space' for the β -decay. The result is a variation in the decay rate depending on the energy of the electron. The spectrum in Fig. 2 shows the resonances in the ^{187}Re decay.

The ripples in the ^{187}Re spectrum are too close together to be resolved in experiments using traditional detectors. The amplitudes of the waves detected by Gatti and colleagues' cryogenic detector are surprisingly high, about 1% of the count rate. At very low energies, the detector becomes noisy and the waves are too close together to be resolved, and at higher energies the waves become negligibly small. But there is a region in between where the ripples in the spectrum are very well resolved (Fig. 2). The shape of each individual ripple can be accurately predicted from the known location of the surrounding atoms. Or the location of the neighbouring atoms can be measured from the shape of the BEFS ripples. The system works like the resonances of bathroom singing: some frequencies (energies) are amplified, and others are extinguished, depending on the shape of the walls.

The beautiful molecular interference pattern measured in the rhenium β -decay demonstrates the limits observed by experimenters in search of the elusive neutrino mass. The molecular influence of the very low-energy β -decay spectra is much bigger than the change in shape caused by a finite electron-neutrino mass. On the other hand, as has happened before, it turns out that a technique developed for fundamental particle research can be used to study the precise



100 YEARS AGO

The present position of the Röntgen rays in military surgery was described by Major J. Battersby in a paper read before the Röntgen Society on Tuesday. ... After the battle at Omdurman 121 British wounded were conveyed to the surgical hospital at Abadie. Of that number there were 21 cases in which the bullet could not be found, or its absence proved by ordinary methods. In 20 out of these 21 cases an accurate diagnosis was arrived at with the help of the rays, the odd case, who was suffering from a severe bullet wound in the lung, being too ill for examination at the time. ... In many cases the X-rays prevented much suffering to the patient, which would have been caused by probing, the use of the finger, or enlarging the wound in the ordinary search for the bullets, as the skiagraph at once indicated the exact position of the bullet. ... With regard to apparatus, the most serious difficulty at present is the best method of generating the primary electrical current for charging the storage batteries. ... In the Sudan a small dynamo, driven by means of a tandem bicycle, answered admirably, and was readily transported by rail and river to Abadie; but as at present constructed, it is unsuitable for mule, camel, or human transport. From *Nature* 12 January 1899.

50 YEARS AGO

Some months ago, Dr. P. Tate gave a review of white-eyed mutants of *Diptera*. In his paper Tate stated that in *Calliphora erythrocephala* the male never manifests the white-eyed character. Some years ago a white-eyed mutant of *Calliphora erythrocephala* appeared in our mass cultures, and in this case both white-eyed females and males occurred. We reared the mutant and obtained a mass culture of it. At that time, however, we were busy with other work, and so we did not investigate the mutant as to the mode of inheritance. Unfortunately, after some time the culture was allowed to die out. It is evident that the mutation found in our culture differs from that of Tate, his mutation being sex-limited, which was not the case with ours. It is interesting to note that the same visible character may be due to different gene mutations. From *Nature* 15 January 1949.

atomic-scale surrounding of an atom. Because cryogenic detectors can be built from many different materials, they may become a tool to study the internal structure of many solids in table-top experiments, without the use of large synchrotron light sources and expensive X-ray monochromators. □

Wolfgang Stoeffl is in N-division, Lawrence Livermore National Laboratory, Livermore,

California 94550, USA.

e-mail: Stoeffl1@llnl.gov

1. Gatti, F., Fontanelli, F., Galeazzi, M., Swift A. M. & Vitale, S. *Nature* **397**, 137–139 (1999).
2. Pauli, W. in *Gesammelte Werke Abt. C. Bd. 1: Physik und Erkenntnistheorie* (ed. Heisenberg, W.) 156–180 (Vieweg, 1984), translated into English by Zacek, G. in *Neutrino Physics* 1–26 (Cambridge Univ. Press, 1991).
3. Fermi, E. *Ricerca Scient.* **2**, 12 (1933).
4. Koonin, S. *Nature* **354**, 486–470 (1991).
5. Conradson, S. D. *Applied Spectrosc.*, **52**, 252A–279A (1998).

Neurobiology

Modulation minimizes masking

Brian C. J. Moore

The ability to distinguish auditory signals from background noises is important for many organisms to detect — and avoid becoming — prey. This ability was probably also significant during evolution, and on page 154 of this issue Nelken *et al.*¹ suggest how the auditory system may have evolved to exploit the properties of natural sounds.

Laboratory studies of auditory masking have traditionally used background sounds (maskers) such as white noise, which contains a range of different frequencies. Such noise fluctuates randomly in intensity, and these fluctuations are independent in different frequency regions. For example, if white noise is passed through two band-pass filters, one centred at 500 Hz and the other at 2,000 Hz, the output of the two filters will show independent, random fluctuations in intensity. More recently, 'comodulated' maskers have been used, in which the fluctuations in intensity are correlated in different frequency regions. These can be created by amplitude modulating a white noise with a low-frequency noise. The low-frequency modulator causes the amplitude of the noise to fluctuate slowly in an irregular way, and the pattern of fluctuation is the same in all frequency regions.

Humans are often much better at detecting signals in comodulated maskers than in white noise, an effect called 'comodulation masking release' (CMR)^{2–4}. However, the extent to which CMR occurs outside the laboratory is unclear. Now, Nelken *et al.*¹ describe a new way to analyse naturally occurring sounds, and show that many such sounds are comodulated. They also show that

the responses of neurons in the cat auditory cortex to a tone signal in noise are greater for comodulated than for unmodulated noise. These findings lead the authors to suggest that the auditory system may have evolved to exploit comodulation in natural situations.

Many properties of auditory masking can be understood in terms of the responses of the basilar membrane within the inner ear (Fig. 1). Each point on this membrane behaves like a filter that responds to a limited range of frequencies. One end is tuned to high frequencies, the other to low frequencies, with a continuous gradation in between. These filters are often called 'auditory filters'. Almost 60 years ago, Fletcher⁵ proposed that, when trying to detect a sinusoidal tone (such as the sound of a tuning fork) in background noise, listeners use the output of a single auditory filter tuned to the frequency of the tone. That filter passes the tone at full intensity, but rejects most of the background noise. Although this theory can account for many aspects of masking⁶, in the mid-1980s Hall *et al.*² showed that, when comodulated maskers were used, the results could be explained only if listeners compared the outputs of auditory filters tuned to different frequencies.

Figure 2 compares the masking produced by white noise (filtered to contain a limited range of frequencies around the signal frequency) with that produced by comodulated noise filtered in a similar way. The threshold for detecting the signal is plotted as a function of the range of frequencies in the noise (the bandwidth). As the bandwidth is increased, the total intensity of the masker increases. For

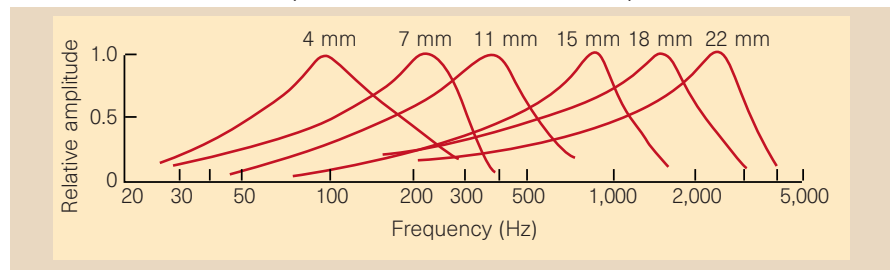


Figure 1 Frequency responses at six different points along the basilar membrane in the inner ear⁸. The membrane is about 35 mm long, and the position of each point is indicated, measured relative to the end that responds best to low frequencies.

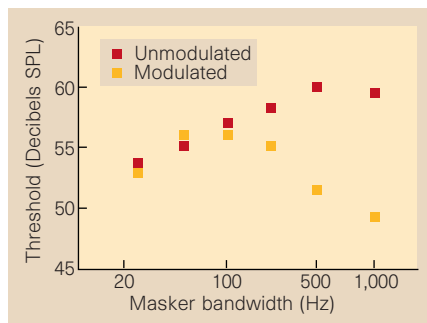


Figure 2 Thresholds for detecting a 1,000-Hz tone as a function of the bandwidth of a noise masker². One masker is unmodulated (random) noise; the other is comodulated noise. For the unmodulated noise, the masked threshold at first increases as the masker bandwidth is increased. This is because more noise passes through the auditory filter centred at the signal frequency. Beyond a certain bandwidth the threshold no longer increases because the added noise falls outside the pass band of the auditory filter centred at the signal frequency. For the comodulated noise, however, the threshold decreases as the bandwidth is increased beyond 100 Hz. SPL, sound pressure level.

unmodulated noise, the masked threshold at first increases with the masker bandwidth, but beyond a certain bandwidth the threshold no longer increases. For the comodulated noise, however, the threshold *decreases* as the bandwidth is increased beyond 100 Hz — in other words, adding more noise to the masker makes the signal easier to hear! Presumably, once the noise bandwidth is large enough, listeners can compare the output of the auditory filter centred at the signal frequency with the outputs of filters tuned away from that frequency. When only the comodulated masker is present, the outputs of all filters fluctuate in a similar way over time. But when the signal is present, the common pattern of fluctuation is disrupted, indicating that the signal is there.

Nelken *et al.*¹ first analysed natural sounds such as animal vocalizations. Each sound was decomposed into a noise-like carrier signal and an envelope representing the amplitude modulation of that signal. (The envelope is like a smooth curve, joining the amplitude peaks of the sound.) To check that this was an appropriate way to analyse the sounds, the authors re-synthesized them from the carrier and the envelope. Where the result was similar to the original, the sounds were described as 'separable'. Mixtures of vocalizations and non-animal sounds were usually separable, whereas the calls of single animals were often non-separable. Nelken *et al.* also developed a 'fluctuation index' by which they quantified the extent to which temporal fluctuations in different frequency bands were correlated over time. About 29% of the sounds showed significant comodulation, and the authors attributed this to two factors: first, animals can synchronize their

calls to form clusters; and second, local turbulence in the atmosphere can affect the propagation of sounds.

The authors then studied neurons in the primary auditory cortex of the cat, and found that the responses were locked to the temporal envelope of modulated noises. Envelope locking often increased with bandwidth. Adding a tone to these sounds disrupted the pattern of envelope locking, providing a cue for detection of the tone. As with humans (Fig. 2), the reduction increased as the bandwidth of the modulated background increased. So could this be the neuronal mechanism that underlies CMR? That remains to be seen. Experiments with humans indicate that CMR involves selectively listening for the signal during the temporal 'dips' in the masker envelope^{3,7}. Also, for human listeners, CMR increases as the masker bandwidth increases over a large range of bandwidths (Fig. 2),

whereas the envelope-locking observed by Nelken *et al.* varied with bandwidth only for the narrower bandwidths. Nonetheless, Nelken and colleagues' results certainly represent progress in the search for the neural mechanisms of CMR. □

Brian C. J. Moore is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

e-mail: bcjm@cus.cam.ac.uk

1. Nelken, I., Rotman, Y. & Yosef, O. B. *Nature* **397**, 154–157 (1999).
2. Hall, J. W., Haggard, M. P. & Fernandes, M. A. *J. Acoust. Soc. Am.* **76**, 50–56 (1984).
3. Hall, J. W., Grose, J. H. & Mendoza, L. in *Hearing* (ed. Moore, B. C. J.) 243–266 (Academic, San Diego, 1995).
4. Moore, B. C. J. *J. Acoust. Soc. Jpn (E)* **13**, 25–37 (1992).
5. Fletcher, H. *Rev. Mod. Phys.* **12**, 47–65 (1940).
6. Patterson, R. D. & Moore, B. C. J. in *Frequency Selectivity in Hearing* (ed. Moore, B. C. J.) 123–177 (Academic, London, 1986).
7. Buus, S., Zhang, L. & Florentine, M. *J. Acoust. Soc. Am.* **99**, 2288–2297 (1996).
8. von Békésy, G. *Experiments in Hearing* (McGraw-Hill, New York, 1960).

Signal transduction

An anchor for activation

Peter ten Dijke and Carl-Henrik Heldin

A critical mechanism for regulating the speed and specificity of signal transduction is to restrict the subcellular localization of the signalling components^{1,2}. Anchor proteins, tethered to specific subcellular structures through protein–protein or protein–lipid interactions, can then serve as a platform to bring them together. Members of the transforming growth factor (TGF)- β family, for example, transmit their signals from the plasma membrane to the nucleus through combinations of serine/threonine kinase receptors and their downstream effectors, known as Smads³.

A report by Tsukazaki *et al.*⁴ in *Cell* now shows how the first intracellular step in the TGF- β /Smad pathway occurs. The authors have identified an anchor protein called SARA (Smad anchor for receptor activation), which recruits Smad2 and Smad3 to the TGF- β receptor. Moreover, SARA mutants that interfere with the subcellular localization of Smad2 potently inhibit TGF- β responses, demonstrating the importance of SARA in the TGF- β signalling pathway (Fig. 1, overleaf).

Smads are pivotal to intracellular signalling by members of the TGF- β family. Smad2 and Smad3 are receptor-regulated Smads that interact with, and become phosphorylated by, activated type-I receptor kinases. The phosphorylated Smad then binds Smad4, and this heteromeric complex is translocated to the nucleus where it controls the transcription of target genes.

Tsukazaki and colleagues⁴ identified SARA by probing a bacterial expression library with Smad2's conserved carboxy-terminal domain. They found that SARA contains a FYVE domain, a motif that has been

identified in 30 other proteins from yeast to mammals^{5,6}, including the mammalian early endosomal antigen (EEA)-1 and hepatocyte growth-factor-regulated tyrosine kinase substrate (Hrs). Hrs, in fact, also seems to be implicated in signalling downstream of serine/threonine kinase receptors (K. Sugamura *et al.*, personal communication). Moreover, the FYVE domain of EEA1 binds the membrane phospholipid phosphatidylinositol-3-phosphate (PtdIns(3)P)^{5,6}, suggesting that FYVE domains may be able to anchor proteins at the inner leaflet of the cell. The FYVE domain could act in a similar way to the pleckstrin homology domain which, in several other signal-transduction molecules, recognizes specific polyphosphoinositides⁷.

Tsukazaki *et al.* show that the subcellular localization of SARA is controlled through interactions mediated by the FYVE domain, probably via membrane-associated PtdIns(3)P. Immunofluorescence confocal microscopy revealed that, whereas SARA is normally present in a punctuate pattern, SARA mutants that lack the FYVE domain are diffusely localized throughout the cell. But although SARA determines the cellular distribution of Smad2, Smad2 has no effect on SARA localization. So, SARA, which is absent from the nucleus, may help to keep non-activated Smad2 in the cytoplasm. Interestingly, SARA and TGF- β type-II receptors localize to common subcellular domains⁸. Importantly, SARA mutants that are mislocalized, but retain their ability to bind Smad2, fail to direct Smad2 to its appropriate subcellular location and strongly inhibit TGF- β -induced transcriptional responses⁴.

The assembly of a heteromeric complex

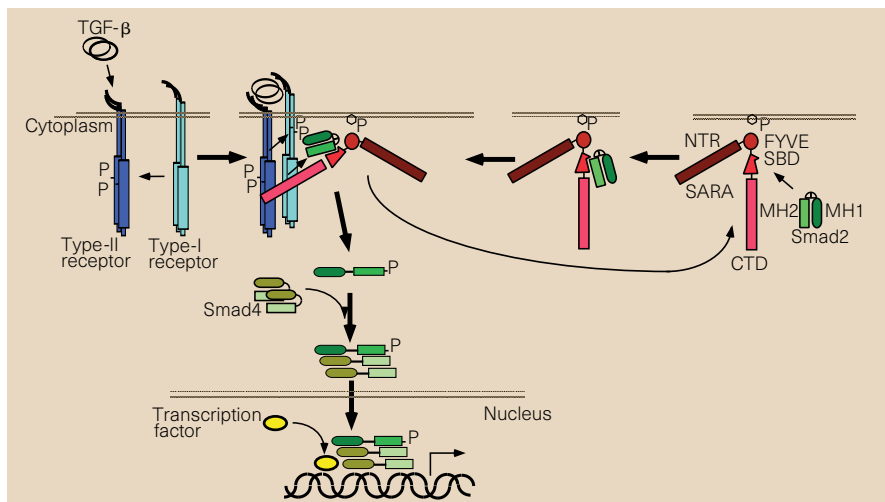


Figure 1 Model for signal transduction in the transforming growth factor (TGF)- β /Smad pathway. Tsukazaki *et al.*⁴ have identified a protein called SARA (Smad anchor for receptor activation) which, through adjacent binding sites for phosphatidylinositol-3-phosphate (PtdIns(3)P) and Smad2, recruits Smad2 to specific regions of the plasma membrane that also contain TGF- β receptors. SARA, Smad2 and the TGF- β -receptor complex cooperatively associate to form a transitional signal-transduction complex. After phosphorylation by a type-I receptor kinase, Smad2 dissociates and reassociates with Smad4 to form a heteromeric complex that translocates to the nucleus and regulates gene transcription. SARA is freed to recruit new Smads for activation by the type-I receptor kinase. NTR, amino-terminal region; SBD, Smad-binding domain; CTD, carboxy-terminal domain; MH, Mad-homology domain.

containing the Smad, SARA and the TGF- β receptor is mediated through direct, cooperative interactions between these components. Smad2 and Smad3 are known^{9,10} to associate with the activated TGF- β type-I receptor, and SARA is now shown to associate with Smad2 through an 85-amino-acid Smad-binding domain, located just on the carboxy-terminal side of the FYVE domain. In addition, SARA interacts through its carboxy-terminal domain with the TGF- β receptor complex, independently of Smad2 binding.

After Smad2 has been phosphorylated by the TGF- β type-I receptor, the components of the complex dissociate — the phosphorylated Smad has a lower affinity for both the receptor⁹ and the Smad-binding domain in SARA. Phosphorylated Smad2 then forms a complex with Smad4, and translocates into the nucleus. SARA and Smad4 interact with the carboxy-terminal domain of Smad2 in a mutually exclusive manner⁴, and once it has been released from the phosphorylated Smad2 molecule, SARA can recruit more non-activated Smad2.

Although SARA interacts with Smad2 and Smad3, Tsukazaki *et al.*⁴ find that it does not bind to Smad1 — one of the receptor-regulated Smads involved in the bone morphogenetic protein (BMP) pathway. Considering the high mechanistic similarity between the TGF- β and BMP signal-transduction pathways, it is likely that Smad1 associates with a SARA-related molecule, and Tsukazaki and co-workers have identified two such candidates. Remaining questions include whether SARA binds directly to type-I or type-II receptors, or whether it binds to receptors other than TGF- β receptors. If SARA does

not bind to activin receptors, it may provide an explanation for the differences in biological responses of TGF- β and activin, which can potentially activate the same Smads. Furthermore, SARA may have a scaffolding function by interacting with other components, such as other receptor substrates.

Tsukazaki *et al.* have shown that SARA is a critical component in TGF- β signalling, recruiting Smad to the TGF- β receptor. SARA may increase the efficiency, as well as the selectivity, of receptor signalling by favouring the phosphorylation of particular substrates and preventing unwanted crosstalk with other pathways. Control of TGF- β signalling may thus involve modulating the levels and subcellular distribution of SARA, or regulating the interactions between SARA and the TGF- β receptor and/or Smad. Finally, SARA — and similar molecules that present substrates — may be useful targets in screening for pharmacological compounds that inhibit TGF- β /Smad signalling. □

Peter ten Dijke and Carl-Henrik Heldin are at the Ludwig Institute for Cancer Research, Box 595, S-751 24 Uppsala, Sweden.

e-mails: Peter.ten_Dijke@LICR.uu.se
C-H.Heldin@LICR.uu.se

1. Pawson, T. & Scott, J. D. *Science* **278**, 2075–2080 (1997).
2. Tsunoda, S., Sierralta, J. & Zuker, C. S. *Curr. Opin. Genet. Dev.* **8**, 419–422 (1998).
3. Massagué, J. *Annu. Rev. Biochem.* **67**, 753–791 (1998).
4. Tsukazaki, T., Chiang, T., Davidson, A., Attisano, L. & Wrana, J. *Cell* **95**, 779–791 (1998).
5. Wiedeman, C. & Cockcroft, S. *Nature* **394**, 426–427 (1998).
6. Corvera, S. & Czech, M. P. *Trends Cell Biol.* **8**, 442–446 (1998).
7. Lemmon, M. A. *et al. Trends Cell Biol.* **7**, 237–242 (1997).
8. Henis, Y. I., Moustakas, A., Lin, H. Y. & Lodish, H. F. *J. Cell Biol.* **126**, 139–154 (1994).
9. Macías-Silva, M. *et al. Cell* **87**, 1215–1224 (1996).
10. Zhang, Y. *et al. Nature* **383**, 168–172 (1996).

Daedalus

A cultured brain

In the past 40 years or so, drug therapy has revolutionized psychiatry. Ideas about complexes, sublimation, super-ego conflicts and so on, have been replaced by concerns for neurotransmitter deficits, endorphins and receptors. But drug therapy is still a crude business. New psychotropic drugs are largely chance products, and their effects and side-effects must be found out the hard way. Daedalus plans to advance the art.

He points out that brain cells can now be cultured *in vitro*. One pioneering study used mouse brain cells, encouraged to grow and divide by epidermal growth factor. (This makes sense; the skin and the brain both grow from the same primitive set of embryonic cells.) In the right environment, says Daedalus, such cultured cells should start to put out dendrites and axons, and make connections with each other. A petri dish bearing a single sheet of cells might be best; their axons could wander freely over the surface of the sheet.

This 'brain pan' of interconnected cells will have no long-range structure or order. But it should be a splendid test-bed for psychotropic drugs, and theories of brain function and dysfunction. Once the brain pan has fully grown, Daedalus will lower test electrodes to touch specific cells in its monolayer. A pulse will be injected into one electrode. The cells around it which fire in response will be identified, and their connections with the central cell will be traced. Then a drug, neurotransmitter or brain metabolite will be added to the culture, and its effect on pulse transmission will be noted.

At first Daedalus will seek to understand the effects of known drugs. He will then test the resulting theories by trying new ones. No test animals will need to be sacrificed or human subjects risked; the most drastic or speculative notions will be easily tried. In particular, he plans to use widely separated electrodes to impose electroshock therapy on the brain pan, and to look for collective epileptic seizures in it.

By damaging selected cells of the brain pan, and noting its responses, Daedalus hopes to gain insights into Parkinsonism, Alzheimer's disease, and the other dismal brain deteriorations. Will the surrounding cells try to make other connections? Will local stem cells differentiate to replace lost neurons? Can a drug, a growth hormone or implanted new cells encourage these processes? Positive answers will bring cheer to ageing citizens and State pension and health agencies alike.

David Jones

Obituary

Alan Hodgkin (1914–98)

Neurophysiologist

Alan Hodgkin, who died on 20 December 1998, was one of the greatest physiologists of the century. He set the foundations for much of modern neuroscience by explaining the origin of the nerve impulse, in an enormously productive collaboration with Andrew Huxley. Their work not only accounted for the explosive ‘firing’ of a nerve impulse, and for conduction of that impulse along the axon (nerve fibre), but it also opened up an entire new science of ‘ion channels’ — the sub-microscopic pores, embedded in the surface membrane of every cell, that regulate the passage of ions and control a multitude of cellular processes.

In July 1939, Hodgkin and Huxley, aged 25 and 21, succeeded in measuring the voltage on the inside of a nerve cell, all previous measurements of nerve activity having been extracellular. Working on squid at Plymouth, Devon, they dissected out the ‘giant axon’ (about 0.5 mm in diameter) and steered a fine glass capillary electrode along its interior. Their remarkable discovery (*Nature* 144, 710–711; 1939) was that during a nerve impulse the intracellular voltage reversed, from its resting level of -60 mV, to reach a peak of about $+40$ mV. This finding destroyed the dogma that had held sway since the turn of the century — that the impulse was produced by a loss of the membrane’s resting ion selectivity — because a collapse of that kind would at most have taken the internal voltage to zero. But within three weeks of the discovery, their research was interrupted by the outbreak of the Second World War.

Most unusually for a physiologist, Hodgkin had taught himself mathematics as an undergraduate, and his adeptness at differential equations was enormously important not only to his later research but also to his wartime work on the development of airborne radar. Microwave communication was an infant technology, yet the spiral-scan and helical-scan centimetre radar sets that he helped develop for fighter aircraft were highly successful in detecting enemy bombers. The ordeals that he underwent, strapped into the tail of fighter aircraft fitted with prototype radar sets, were both thrilling and testing — not least because the 15 kV supply was prone to arc over in the rarefied atmosphere at 15,000 feet, occasionally setting fire to the equipment.

Far from being purely mathematical, Hodgkin’s work was based on cutting-edge



technology. When, from 1946, he was able to return to research on nerve, he built state-of-the-art electronic equipment — although modern-day electrophysiologists might be amazed that home-made cathode followers and vacuum-tube feedback amplifiers could have been exploited so successfully. One breakthrough was measurement of the current through the nerve membrane when its voltage was ‘clamped’ to different levels (thereby disrupting the positive feedback loop that generated the explosive rise in voltage).

But perhaps his greatest contribution was to invent a way of accounting for the measurements, through a set of differential equations representing separate sodium- and potassium-selective pathways. The results and analysis of those experiments were published in a series of five remarkable papers by Hodgkin, Huxley and Bernard Katz in 1952. The culmination, a quantitative description of the events underlying nerve activation (*J. Physiol.* 117, 500–544; 1952), remains a masterpiece. Hodgkin’s personal account of the research, and of his wartime exploits, is told in *Chance and Design* (Cambridge Univ. Press, 1992).

That ‘classical’ description of the voltage-sensitive opening and closing of sodium- and potassium-selective conductances opened the way for modern molecular approaches. It predicted the existence of discrete ion channels for Na^+ and K^+ , and it accurately specified their kinetic properties; it forecast their tetrameric structure, with four separate voltage-sensors; and it predicted the occurrence of ‘gating currents’ needed to open the channels. For more than 40 years,

that work has been the standard against which subsequent ‘patch clamp’ and molecular cloning approaches have been evaluated.

Hodgkin’s research was remarkable not only for its mathematical elegance, but also for the sheer beauty of his experimental results. He totally immersed himself in his experiments and their analysis, and to this end he always kept his research group very small, with just one or two collaborators. For these most fortunate colleagues the experience was richly rewarding, but demanding. He always took great delight in unexpected experimental findings, and in the intellectual exercise of accounting for the reality of nature.

In 1963 Alan Hodgkin was awarded the Nobel Prize for Physiology or Medicine, with Andrew Huxley and Jack Eccles; he was knighted in 1972, and the following year was appointed to the Order of Merit (a British honour restricted to just 24 people at any one time). From 1970 to 1975 he served as President of the Royal Society, and from 1978 to 1984 as Master of Trinity College, Cambridge, where (since 1932) he had been first an undergraduate and subsequently a fellow. Under his leadership the Royal Society became more open in its formulation of science policy, and made preparations to meet the major changes in science funding that were looming.

In 1970, Hodgkin turned his attention to the visual system, when the exciting new technique of intracellular recording from retinal neurons was brought to his laboratory by Denis Baylor. Hodgkin transformed the study of rod and cone photoreceptors, bringing to the subject the incisiveness and quantitative rigour that were the hallmark of his research on nerve and muscle. Our understanding of transduction in retinal photoreceptors is based on this work, carried out between 1970 and 1988.

For more than half a century Alan’s career was supported by his wife Marni — herself the daughter of a Nobel laureate, Peyton Rous. She well understood the demands of scientific research, and revelled in the entertaining associated with his prestigious posts; together they shared a wide-ranging love of the arts.

Alan Hodgkin’s wonderful approach to biology stimulated generations of neuroscientists to the study of nerve, muscle and photoreceptors. His death marks the end of a classic era in neurophysiology.

Trevor Lamb

Trevor Lamb is in the Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK.
e-mail: TDL1@cam.ac.uk

Lizards took express train to Polynesia

The pattern of human colonization of the islands of the central and eastern Pacific is contentious. Two models have been widely considered: the 'express train to Polynesia' and the 'entangled bank' hypotheses¹⁻⁴. Here I analyse the mitochondrial DNA sequences of the lizard *Lipinia noctua*, which lives alongside humans on these Pacific islands, with a view to distinguishing between these two hypotheses. From a phylogenetic analysis of mitochondrial DNA sequence variation, I find that these lizards colonized the central and eastern Pacific as a result of human-mediated dispersal, presumably as stowaways on early Polynesian canoes. The extreme genetic similarity between the different colonies indicates rapid colonization from a single source, which I take as support of the express-train hypothesis.

The Pacific Ocean is the world's largest geographic feature, encompassing more area than all the continents combined, and the islands of the eastern Pacific were the last part of the world to be colonized by humans. The voyage to these islands has attracted attention because it could provide clues as to how human cultures evolved. The entangled-bank hypothesis² argues that movement into the central and eastern Pacific was a gradual event, occurring over an extended period from different Melanesian populations. In contrast, the express-train hypothesis¹ argues for rapid colonization from southeast Asia, with colonists moving through Melanesia and into the Pacific with little genetic exchange occurring between different groups.

Archaeological⁵, linguistic⁶ and genetic⁷⁻⁹ data indicate that human migration from the Taiwan region to the Bismarck archipelago, northeast of New Guinea, occurred between 3500 BC and 1600 BC. The Lapita expansion, from the Bismarck islands into the central Pacific, is thought to have happened over a short period, perhaps only a few hundred years^{1,3,5}. The further extension of humans into the eastern Pacific was slower, and the most remote regions were not reached until much later⁴ (Fig. 1).

Lipinia noctua, a small lizard native to New Guinea, is a single species with one of the broadest distributions of any lizard. It occurs from the Papuan region throughout Oceania, to the Hawaiian islands in the northeast, and Easter Island and Pitcairn Island in the southeast. Because of the large ocean barriers and the morphological similarity between island populations, dispersal of the lizard from New Guinea was presumed to have been mediated by humans.

To investigate any potential congruent phylogeographic patterns between humans and this lizard, I collected mitochondrial

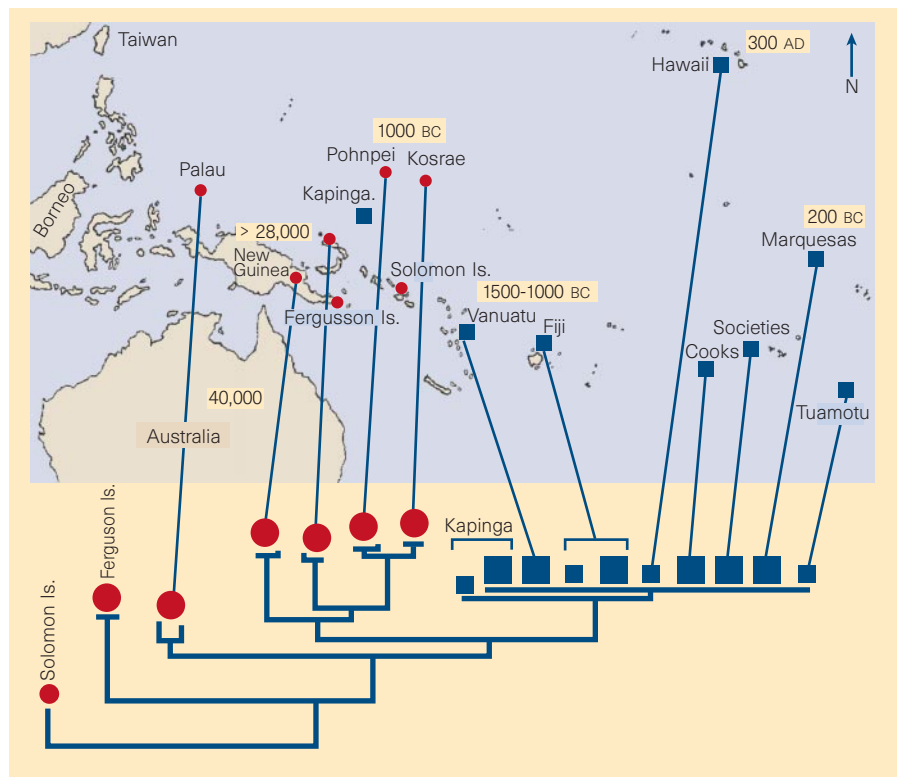


Figure 1 Maximum-parsimony phylogram for *Lipinia noctua*. Maximum-likelihood analyses produced the same topology. Symbol size represents sample size (small, $n=1$; large, $n=2$). Localities denoted by blue circles are all genetically distinct from one another (mean sequence divergence, 9.7%) and represent natural pre-human dispersal¹⁰. Localities denoted by red squares are genetically similar (mean sequence divergence, 0.008%) and represent human-mediated dispersal within the past 4,000 years. Kapinga indicates Kapingamarangi atoll. Dates represent approximate time of first human settlement⁴. (R. Fisher, J. Lum and D. Mindell, personal communication.)

DNA samples from 29 *L. noctua* lizards from 15 island populations (Fig. 1). From these samples, I sequenced and aligned a 300-base-pair portion of the gene encoding cytochrome *b* and analysed the results by maximum-parsimony and maximum-likelihood phylogenetic reconstruction methods using appropriate outgroups (Fig. 1).

The entangled-bank hypothesis predicts that *L. noctua* populations in the central and eastern Pacific should be paraphyletic (red and blue populations intermixed in Fig. 1) as a result of several dispersal events that reflect the different human migrations from Melanesia. The express-train hypothesis, however, predicts that *L. noctua* in the central and eastern Pacific should be monophyletic (red populations grouped together in Fig. 1) and show a high degree of genetic similarity because of a bottleneck associated with rapid human-mediated dispersal.

The colonization pattern for *L. noctua* matches the predictions of the express-train hypothesis (Fig. 1). All central and eastern Pacific populations of *L. noctua* form a monophyletic clade that differ by no more than two base-pair substitutions. As

humans moved rapidly into the central Pacific, they could have inadvertently carried with them in their canoes a few lizards (or perhaps a single gravid female), causing the observed founder effect.

Although they are geographically part of Micronesia, the people of Kapingamarangi atoll are Polynesian in origin^{4,6}. The *L. noctua* from there are also of the central/eastern clade, which strengthens the association between *L. noctua* and human colonization. The lizard populations from the central and eastern Pacific should be equally closely related to an original source population, presumably somewhere in the Bismarck or Solomon archipelagoes. However, this source population, identifiable by greater haplotype diversity, has not yet been sampled. When it is identified, it may reveal archaeological information about the early stages of the express train's journey. The populations of *L. noctua* in Micronesia show a large degree of genetic divergence, indicating that they are the result of pre-human natural dispersal¹⁰.

The application of comparative phylogeography to evolutionary history, together

with information about population structure from other such 'vagabond' fauna and flora, should continue to shed light on the vagaries of human evolution.

Christopher C. Austin

*Evolutionary Biology Unit,
South Australian Museum,
Adelaide 5000, Australia*

*Present address: Institute of Statistical Mathematics,
4-6-7 Minami-Azabu Minato-ku,
Tokyo 106-8569, Japan
e-mail: caustin@ism.ac.jp*

1. Diamond, J. M. *Nature* **336**, 307–308 (1988).
2. Terrell, J. *Antiquity* **62**, 642–657 (1988).
3. Clegg, J. B. *Lancet* **344**, 1070–1071 (1994).
4. Bellwood, P. S. in *The Colonization of the Pacific: Some Current Hypotheses* (eds Hill, A. V. S. & Serjeantson, S. W.) 1–59 (Oxford Univ. Press, 1989).
5. Kirch, P. V. & Hunt, R. L. *Radiocarbon* **30**, 161–169 (1988).
6. Pawley, A. & Green, R. *Ocean Ling.* **12**, 1–67 (1973).
7. Hagelberg, E. & Clegg, J. B. *Proc. R. Soc. Lond. B* **252**, 163–170 (1993).
8. Redd, A. J. *et al. Mol. Biol. Evol.* **12**, 604–615 (1995).
9. Martinson, J. J. in *Molecular Biology and Human Diversity* (eds Boyce, A. J. & Mascie-Taylor, C. G. N.) 171–195 (Cambridge Univ. Press, 1996).
10. Thorpe, R. S., McGregor, D. P., Cumming, A. M. & Jordan, W. C. *Evolution* **48**, 230–240 (1994).

The first true inorganic fullerenes?

Boron nitride and materials of composition MX_2 , where M is molybdenum or tungsten and X is sulphur or selenium, can form fullerene-like structures such as nested polyhedra or nanotubes^{1–3}. However, the analogy to the carbon fullerene family⁴ falls short because no small preferred structure akin to C_{60} (ref. 5) has been found. We have discovered nano-octahedra of MoS_2 of discrete sizes in soots that we prepared by laser ablation of pressed MoS_2 targets. These nano-octahedra are much larger than C_{60} structures, having edge lengths of about 4.0 and 5.0 nanometres, and may represent the first 'inorganic fullerenes'.

Targets were prepared by pressing 98% pure MoS_2 powder and ablated using a KrF pulsed excimer laser (8 Hz, 248 nm, ~300 mJ per pulse, ~20 ns per pulse, ~10 J cm⁻²) under flowing helium or argon (500–800 torr, ~90 cm³ min⁻¹). The beam was moved every 4 minutes during the 20-minute runs to strike fresh target material, with the chamber and target temperature ranging from 30 to 600 °C. The soot generated was collected, ultrasonicated in acetone, and applied to a grid for imaging by transmission electron microscopy (TEM).

Soot produced between 30 and 500 °C contained crystalline and amorphous MoS_2 fractions, as well as smaller rhomboidal, rectangular and hexagonal structures 3 to 5 nm long with two or three layers. The crystalline material included large sheets and tubes and a variety of nested polyhedra 15

to 35 nm long that were similar to those produced previously⁶. Above 550 °C, only crystalline folded sheets of MoS_2 were produced.

TEM stage-tilting experiments on 30 two- and three-layered structures showed that the small rhomboids, rectangles and hexagons were different projections of the same three-dimensional structure: an octahedron (Fig. 1a). The TEM image for a closed three-layer structure changes with tilts of 10° and 20° (Fig. 1b). The image at 0° is the projection expected for an octahedron orientated such that only two triangular faces are seen. When it is tilted, the projection changes, resulting in a nearly rectangular projection at 20°. Stick models depicting how an octahedron's projection changes with tilting are also shown in Fig. 1b. The model octahedron was orientated to project a match to the 0° image, and the model was then tilted with the same axis used in the TEM. Many other TEM tilt

sequences could also be generated with projections of a model octahedron. In some cases, slight movements of the particles on the TEM grid ruined the correlation, but individual images could still be represented by the projection of an octahedron.

The edge length of the octahedron may be calculated from TEM projections, assuming a regular octahedral structure. A histogram of edge lengths for 30 different structures is shown in Fig. 1c. Two pronounced peaks are seen at 12–13 and 16 times the *a* lattice constant (the Mo–Mo distance, 3.16 Å) of MoS_2 for two- and three-layer species, respectively. The spacing between the layers is about 0.6 nm, in good agreement with the interlayer spacing in bulk MoS_2 . The edge of the three-layer species is about four *a* lattice constants larger than that of the underlying two-layer structure, exactly the size required to maintain the bulk interlayer spacing.

Although the reasons for these specific sizes are not clear, a preference for two- and three-layer structures may be associated with the two- and three-layer polytypes⁷. The octahedral shape might be anticipated for a closed MoS_2 structure as the triangular faces share the symmetry of the trigonal Mo and S sublattices. Rounded corners and edges are also expected for MoS_2 sheets, which cannot be severely bent without strain. Energy-dispersive spectroscopy indicated a Mo:S ratio of about 1:2 with no detectable impurities. Satisfying such a ratio exactly is impossible in an octahedron, but several arrangements come close. For example, the Mo–S coordination could remain trigonal prismatic, as in the bulk form, with a given face being slightly rich or poor in sulphur. The structure at the vertices is unclear, but either a four-membered Mo ring³ or a single Mo atom might be stable (B. Parkinson, personal communication).

TEM measurements could be performed only on nano-octahedra that were separated from the agglomerates formed on the TEM grid. Consequently, we cannot yet estimate the density of nano-octahedra in the laser-generated soots. We are purifying these inorganic fullerenes so that we can ascertain their properties, and are also finding out whether similar cage structures can be made using other layered materials.

P. A. Parilla, A. C. Dillon, K. M. Jones, G. Riker, D. L. Schulz, D. S. Ginley, M. J. Heben

*National Renewable Energy Laboratory,
1617 Cole Boulevard, Golden,
Colorado 80401-3393, USA*

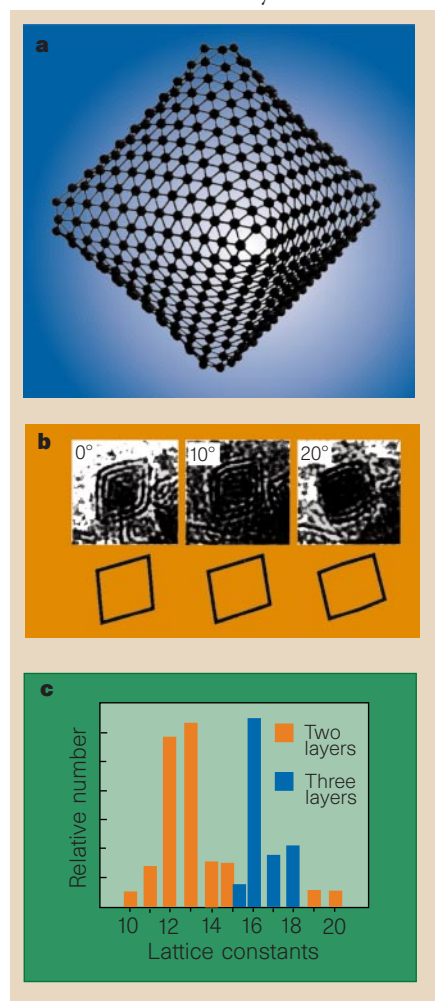


Figure 1 Structure of the molecules. **a**, Model octahedron with an edge length of 12 *a* lattice constants showing only the molybdenum sublattice. **b**, Transmission electron microscope images and model-generated projections for a three-layer MoS_2 rhomboid at 0° and undergoing tilts of 10° and 20°. **c**, Histogram of octahedral edge lengths in *a* lattice constants determined from 30 rhomboidal, rectangular and hexagonal projections observed by TEM.

1. Tenne, R., Margulis, L., Genut, M. & Hodes, G. *Nature* **360**, 444–446 (1992).
2. Chopra, N. G. *et al. Science* **269**, 966–967 (1995).
3. Tenne, R. *Adv. Mater.* **7**, 965–995 (1995).
4. Kroto, H. *Science* **242**, 1139–1145 (1988).
5. Kroto, H. W. *et al. Nature* **318**, 162–163 (1985).
6. Feldman, Y., Wasserman, E., Srolovitz, D. J. & Tenne, R. *Science* **267**, 222–225 (1995).
7. Wilson, J. A. & Yoffe, A. D. *Adv. Phys.* **18**, 193–335 (1969).

Length of tail streamers in barn swallows

Both natural and sexual selection affect the expression of secondary sexual characters¹, so any secondary sexual character will be affected by a mixture of selection pressures. Evans² has compared the flight performance of male barn swallows (*Hirundo rustica*) whose tail lengths have been altered with that of controls, and concludes that the outermost tail feathers of these birds are mainly the outcome of natural selection.

If tail length has been determined by natural selection, then both elongation and shortening of tail feathers should lead to reduced flight performance. If sexual selection is involved, it will reduce the flight performance of males with elongated tails, so shortening the tail will improve performance. It is widely accepted that sexual selection cannot on its own be responsible for the elongated tail, but it may have an influence, as males have allometrically longer tails and streamers than females³.

The average difference in tail length between the sexes is about 15 mm in northern European barn swallows⁴. Because the female tail presumably represents the aerodynamic optimum in adult barn swallows^{3,4}, shortening the male tail by less than 15 mm should provide a test for the sexual selection hypothesis (Fig. 1). But Evans shortened male tails by 20 mm, reducing them beyond the aerodynamic optimum. His experiment could therefore demonstrate reduced flight performance only in support of natural selection, and so could not discriminate between the two hypotheses.

Norberg⁵ has proposed an aerodynamic mechanism to account for the elongated tail streamers in barn swallows, but Evans did not test this. According to Norberg, when the tail is lowered, the tail streamer is bent by air flow, causing an aeroelastic twist through the entire feather that is passively rotated. This rotation turns the outer tail into a flap that provides increased lift of the entire tail. For this to work, the tail must be considered as a finely tuned structure in which the entire tail is a coadapted complex of characteristics. If the tail streamers are manipulated beyond their aerodynamic optimum, other characteristics will also be affected, such as planform, curvature of the feather shaft, flexural stiffness, and torsional rigidity of the outermost tail feathers, which supposedly evolved to compensate for increased pitching moments caused by exaggerated streamers⁵. Any experimental shortening is therefore liable to disrupt performance even in male swallows.

The tail of male barn swallows has some functional utility, but it is the difference

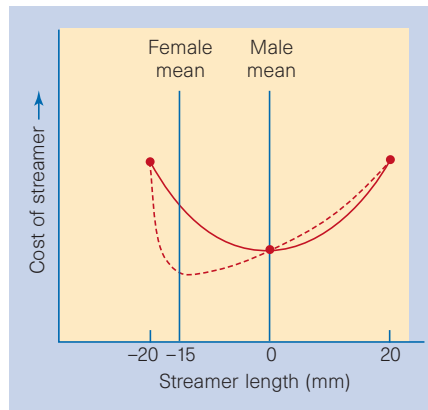


Figure 1 Models for the relation between streamer length and aerodynamic cost in barn swallows, as tested in tail-manipulation experiments. Filled circles, tail-streamer manipulations and effects described by Evans²; the solid curve shows his model. Our alternative model (broken curve) assumes that female streamer length represents the aerodynamic optimum and associated costs away from this optimum. If male streamer length has been modified by sexual selection, this gives an increased cost in relation to female streamer length. We have assumed that the differences between the sexes are controlled for allometric size differences.

between the sexes that is crucial to mate choice. Longer tails in male barn swallows may increase several components of sexual selection (reviewed in refs 4, 6), which has been demonstrated by manipulating apparent tail length without affecting aerodynamic function⁷. All the evidence indicates that the difference in tail length between the sexes is caused by sexual selection.

Anders Hedenström*, Anders Pape Møller†

*Department of Animal Ecology,
Ecology Building, S-223 62 Lund, Sweden
e-mail: anders.hedenstrom@zooekol.lu.se

†Laboratoire d'Ecologie, CNRS URA 258,
Université Pierre et Marie Curie,
Case 237, F-75252 Paris, Cedex 05, France

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Evans, M. R. *Nature* **394**, 233–234 (1998).
3. Hedenström, A. *Trends Ecol. Evol.* **10**, 140–141 (1995).
4. Møller, A. P. *Sexual Selection and the Barn Swallow* (Oxford Univ. Press, 1994).
5. Norberg, R. Å. *Proc. R. Soc. Lond. B* **257**, 227–233 (1994).
6. Møller, A. P. et al. *Proc. R. Soc. Lond. B* **265**, 409–414 (1998).
7. Møller, A. P. *Behav. Ecol. Sociobiol.* **32**, 371–376 (1993).

Evans replies — I have demonstrated that the tail streamers of male swallows are mainly the product of natural selection, although I suggested that sexual selection may have extended the streamer by less than the length the birds' tails were shortened in my experiments (20 mm)¹. My objective was to test whether tail streamers evolved as a result of sexual selection or of natural selection. Over the past eleven years, Møller and co-workers have shown that males with longer tails have many mating advantages². They conclude that “tail

ornaments in the monogamous swallow have evolved as a result of female choice”³.

In the population of barn swallows I studied, the difference in male and female streamer length is 14 mm, although there is extensive overlap between the sexes, so the manipulation of 20 mm I used is greater than the degree of sexual dimorphism. Hedenström and Møller say this invalidates experiments on sexual dimorphism. I chose 20 mm as this was the ‘standard size’ used by Møller and co-workers in their studies of sexual dimorphism in this species.

Hedenström and Møller claim that female streamer length represents the aerodynamic optimum in swallows. Yet in a previous study⁴, Møller and colleagues examined the function of the tail streamer in female swallows and concluded that female streamer length was longer than the optimum, possibly because of a genetic correlation with male streamer length. They suggested that the short tail streamers of juvenile swallows “must be close to the optimum under natural selection”⁴. It follows that female streamer length is unlikely to be the naturally selected optimum, as Hedenström and Møller have assumed⁵. Møller has also shown that male and female swallows have different morphologies⁶, so female streamer length would not represent the aerodynamic optimum for males even if female streamers were naturally selected.

My previous experiment could be improved by making a series of small removal manipulations, when the line relating flight performance to streamer length should have a turning point at the boundary between the naturally and sexually selected regions⁵. The results of such an experiment^{6–8} indicate that for the independent flight parameters there is a minimum 12 mm below the original streamer length. They also show that female tail streamers have been extended by a similar amount beyond the aerodynamic optimum, and so are in accord with Møller's earlier results⁴. To treat female streamer length as the aerodynamic optimum is therefore misleading.

Finally, my experiments did not attempt to test any particular aerodynamic mechanism². I have tested various functional hypotheses for streamer evolution, but the mechanism underlying this functional relation has not yet been proven. Our results indicate that the streamers of both male and female swallows evolved primarily as a result of natural selection, but that they have been extended by about 12 mm beyond this point by sexual selection.

Matthew R. Evans

Department of Biological Sciences,
University of Stirling, Stirling FK9 4LA, UK
e-mail: mre2@stir.ac.uk

1. Evans, M. R. *Nature* **394**, 233–234 (1998).
2. Møller, A. P. *Sexual Selection and the Barn Swallow* (Oxford Univ. Press, 1994).

3. Møller, A. P. *Nature* **332**, 640–642 (1988)
4. Cuervo, J. J., de Lope, F. & Møller, A. P. *Behav. Ecol.* **7**, 132–136 (1996).
5. Evans, M. R. & Thomas, A. L. R. *Proc. R. Soc. Lond. B* **264**, 211–217 (1997).
6. Buchanan, K. L. & Evans, M. R. *Behav. Ecol.* (submitted).
7. Rowe, L. V., Buchanan, K. L. & Evans, M. R. *Proc. R. Soc. Lond. B* (submitted).
8. Evans, M. R., Buchanan, K. L. & Rowe, L. V. *Proc. Natl Acad. Sci. USA* (submitted).

Sign language 'heard' in the auditory cortex

The upper regions of the brain's temporal lobe are important both for hearing and for comprehending spoken language. We have discovered that these regions can be activated by sign language in congenitally deaf subjects, even though the temporal lobe normally functions as an auditory area. This finding indicates that, in deaf people, the brain region usually reserved for hearing may be activated by other sensory modalities, providing striking evidence of neural plasticity.

The auditory areas consist of the primary auditory cortex and the auditory association area (the supratemporal gyrus). The neural network that projects from the inner ear to the primary auditory cerebral cortex is formed without any auditory input, whereas post-processing neurons develop by learning with proper neural input. The learning period for the mother tongue is thought to be below five to six years of age¹. Reducing the auditory signals during the critical language-learning period can severely limit a child's potential for developing an effective communication system². 'Pre-lingual deaf' patients, who were deafened before acquiring language, communicate using sign language.

In an attempt to understand how these auditory areas function in the congenitally deaf, we used positron emission tomography (PET) to measure cortical activation during a sign-language task. In the main experiment we sought to localize the 'sign

language' areas, but a secondary experiment was set up to localize both the auditory areas that had been dormant and the visual areas.

In the main experiment, the subject viewed a video of sign-language words being signed by a native signer; a still frame of the video was viewed in the control task. PET images were seen using statistical parametric mapping software³, and maps were superimposed onto magnetic resonance images of the subject's brain for spatial localization. We found that sign language activated the supratemporal gyri bilaterally (left, $z = 4.52$; $P = 0.005$, corrected; Fig. 1).

The subject was scheduled to have a cochlear implant in his left ear. The implant is an artificial prosthesis, inserted into the inner ear, that electrically stimulates the cochlear nerve and enables the profoundly deaf to hear sounds. To distinguish the supratemporal gyri from the visual and dormant auditory areas, a secondary experiment was performed after the operation, consisting of an auditory task, a visual task and rest. In the visual task, the subject watched a video showing someone moving both hands up and down in a meaningless manner. In the auditory task, the words of the tape were delivered through the cochlear implant. The visual stimulation was found to activate the visual cortex in the occipital lobe ($P = 0.001$, corrected), and the auditory stimulation activated the right primary auditory cortex, contralateral to the auditory input ($P = 0.002$, uncorrected) (Fig. 1).

Pre-lingual deaf people can hear when a cochlear implant is switched on, but this does not allow them to understand words. Language stimulation through the implant activates only the primary auditory cortex in the pre-lingual deaf, whereas in the post-lingual deaf it activates both the primary and the secondary auditory areas⁴. The result of our secondary experiment was compatible with these findings. Our study of native signers and those who learnt sign language later showed that the nature and timing of sensory and linguistic experience

significantly affect the development of the language systems of the brain⁵.

In bilingual subjects (those with both signed and spoken language), sign language activates the visual areas⁶, whereas our study showed activation of the auditory area in the sign-language task. Because our subject had never received auditory input while the neural network was being formed, it seems that the supratemporal lobe was engaged in processing sign language. Using sign language elicits considerable activation of the left hemisphere in Broca's area and Wernicke's area, as well as of the right hemisphere⁷, whereas our results indicated limited activation of Wernicke's area by sign-language words.

This cross-modal plasticity is also seen in visual areas. Braille-reading blind subjects have activation of the primary and secondary visual cortical areas when they perform tactile tasks⁸, although congenitally blind Braille readers have activation of visual reading areas but not primary visual cortex⁹. Our results indicate that the primary auditory cortex of deaf people is reserved for hearing sounds, whereas the secondary areas are used for processing sign language. This cross-modal non-plasticity of the primary auditory cortex is supported by functional magnetic resonance imaging of a congenitally deaf subject¹⁰, which suggests that the primary projection areas might be rigidly organized.

We observed that sign language activates the 'language' areas but not primary auditory cortex. The finding that, after a cochlear implant is in place, spoken words activate primary auditory cortex but not adjacent language areas indicates that primary auditory cortex still functions as an auditory area in this patient. We also identified the 'sign-language' area as the supratemporal gyri, which is usually the auditory area.

Hiroshi Nishimura*†, Kazuo Hashikawa‡, Katsumi Doi†, Takako Iwaki†, Yoshiyuki Watanabe§, Hideo Kusuoka*, Tsunehiko Nishimura*, Takeshi Kubo†

*Division of Tracer Kinetics, †Department of Otorhinolaryngology, ‡Department of Nuclear Medicine, and §Department of Radiology, Osaka University Medical School, Yamada-oka, Suita City, Osaka, Japan
e-mail: hiroshi@tracer.med.osaka-u.ac.jp

1. Osberger, M. J. et al. *Ann. Otol. Rhinol. Laryngol.* **100**, 883–888 (1991).
2. Fitch, J. L., Williams, T. F. & Etienne, J. E. *J. Speech Hear. Disord.* **47**, 373–375 (1982).
3. SPM96 (Wellcome Department of Cognitive Neurology, London, 1996).
4. Naito, Y. et al. *Acta Otolaryngol.* **117**, 490–496 (1997).
5. Neville, H. J. et al. *Brain Lang.* **57**, 285–308 (1997).
6. Soderfeldt, B. et al. *Neurology* **49**, 82–87 (1997).
7. Neville, H. J. et al. *Proc. Natl Acad. Sci. USA* **95**, 922–929 (1998).
8. Sadato, N. et al. *Nature* **380**, 526–528 (1996).
9. Büchel, C., Price, C., Frachowiak, R. S. J. & Friston, K. *Brain* **121**, 409–419 (1998).
10. Hickok, G. et al. *Hum. Brain Mapping* **5**, 437–444 (1997).

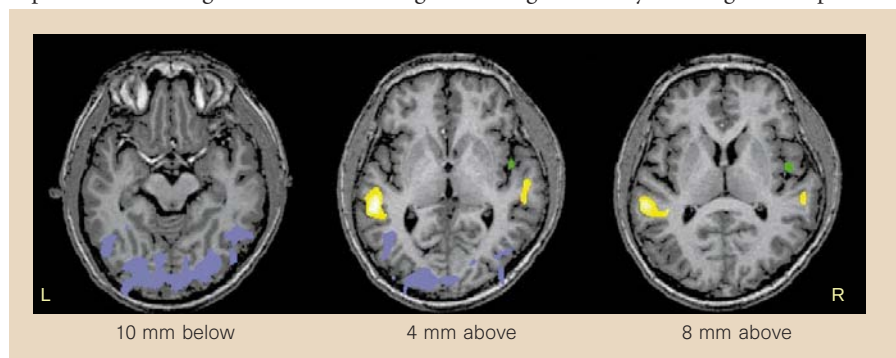


Figure 1 Activation of areas of the brain. The activated areas were superimposed onto the three horizontal sections (10 mm below, and 4 mm and 8 mm above, the intercommissural plane) of the subject's magnetic resonance image. Yellow areas were activated by sign language in the main experiment; green areas were activated by audition, and blue areas by vision, in the secondary experiment.

Thinking about consciousness

Consciousness and Human Identity

edited by John Cornwell
Oxford University Press: 1998. 256 pp.
£19.99, \$25

Paul E. Griffiths

Scientists should not seek to determine the nature of consciousness at all. They should take consciousness for granted and study its correlates in their various domains of enquiry. So, rather surprisingly, says the philosopher of science Peter Lipton in the final contribution to *Consciousness and Human Identity*. Lipton argues against an approach to consciousness that tries to integrate science and philosophy. This is even more surprising given that such integration was the aim of the conference from which the essays in this volume are drawn, the second conference of the Science and the Human Dimension Project at Jesus College, Cambridge. The question of whether consciousness is anything more than the physical processes that scientists discover, says Lipton, is best handled by "leaving the identity question to torment philosophers".

Although Lipton is the only explicit advocate of this view, several other contributors to the volume propose that the problem of consciousness is a conceptual one: the problem is intractable because the way we think about consciousness is not readily commensurable with the modes of thought in the sciences from which we seek explanation. If this is correct, then until a conceptual revolution comes along, attempts to integrate the vernacular understanding of the mind with our growing understanding of the brain may be futile. Since the best source of conceptual revolutions is the struggle to solve particular empirical problems, compartmentalized enquiry may, paradoxically, be the best way to achieve conceptual integration.

The problem of consciousness is compared several times in this volume to the problem of vitalism in nineteenth-century biology. The full extent of the parallel, however, is not explored. In the early nineteenth century, biology contained an array of positions strikingly comparable to those in consciousness studies today. Leading figures such as Karl Ernst von Baer held a Kantian view of the nature of living things akin to the 'new mysterion' view of consciousness held by Colin McGinn and others.

Kantian biologists believed that organisms were entirely material and totally governed by the laws of physics. However, they insisted that it was beyond the capacity of the human intellect to understand organized growth in these terms. The limitations of the intellect required that organisms be understood as intrinsically goal-directed systems



JERZY MAREK/BRIDGEMAN ART LIBRARY

Out for the count: consciousness may be defined as those states that typically begin when we wake and continue through the day until we fall asleep, die or otherwise become 'unconscious'.

whose evolution (in its original sense) was to be understood as a process intrinsically directed towards the adult form, not as the result of the operation of physical and chemical laws.

As Timothy Lenoir has argued, it was these Kantian biologists, rather than contemporary gung-ho materialists, who created the descriptive embryology that was the great achievement of the period. An empirical vindication of materialism was quite impossible without the concept of the organism as being represented by the cell theory and by the physics of heat with the consequent rigorous formulation of the conservation laws, or without the techniques of manipulative, experimental embryology, all of which were yet to appear. In the eighteenth and early nineteenth century, attempts to link embryology to physics and chemistry led to flimsy 'how possibly' stories rather than productive research.

A further lesson that might be drawn from the case of vitalism is that the 'problem'

of consciousness may not be the sort of problem that gets 'solved'. In this book, John Searle gives what he calls a "common-sense definition" of consciousness: "those states of sentience or awareness that typically begin when we wake from a dreamless sleep and continue through the day until we fall asleep again, die or go into a coma, or otherwise become 'unconscious'". Searle's aim is to show that we know what we are about in consciousness studies although we lack even the vaguest sketch of a theory of consciousness.

Searle is quite correct on this point, but he cannot rule out the possibility that, as we investigate the subject, we will transform the conceptual landscape to the point where the initial definition is not explicated, but rather is left behind as an episode in intellectual history.

The vital forces of the eighteenth century, the *vis essentialis* of C. F. Wolff and the *bildungstrieb* of G. F. Blumenbach, were intended to denote the unknown factors that

cause organized growth and make offspring conform to the type of their parents. Were the *bildungstrieb* and its successors reduced to the physical, or were they eliminated as the false posits of an erroneous theory? Neither description is useful. These constructs certainly fell into disrepute among late nineteenth-century embryologists, but the gene-centred accounts of growth and heredity that are now conventional are closer to the views of von Baer than to those of the later, materialist proponents of 'developmental mechanics'.

The debate over vitalism has become an episode in intellectual history that can only be adequately understood in the light of the conceptual possibilities and the potential for practical research that were available in the early nineteenth century. Similarly, I suspect that we will neither 'solve' the problem of consciousness, nor 'eliminate' consciousness in favour of some superior and incommensurable way of conceptualizing the same phenomenon. Rather, our successors will look back across an intellectual gulf, parodying our conception of consciousness in their undergraduate textbooks and being corrected by more sensitive professional historians.

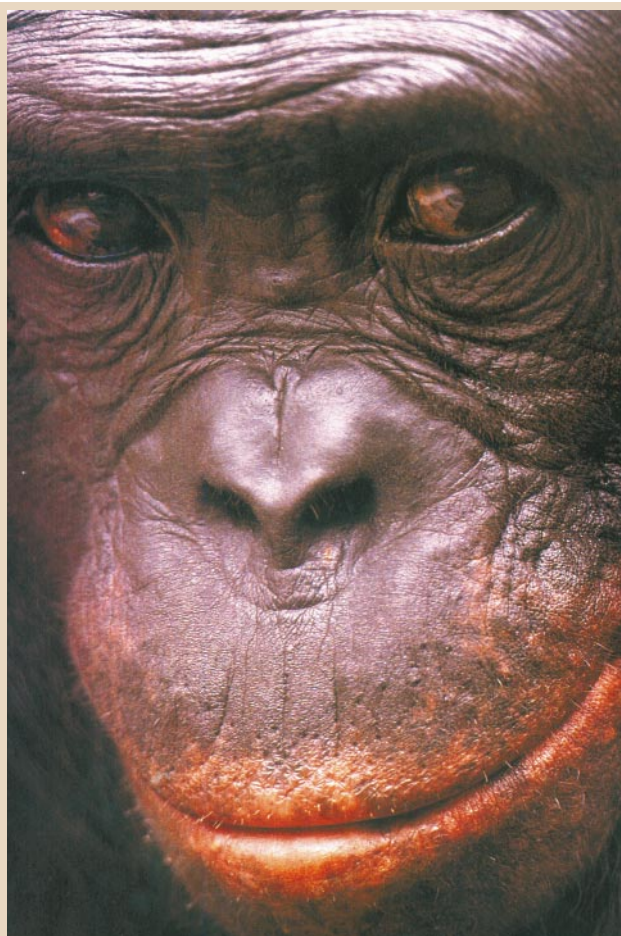
John Cornwell is to be congratulated on the editing of this collection, which forms an effective introduction to consciousness studies. The papers have been rewritten since the conference at which they were presented to form a more coherent whole, and make frequent reference to one another. Cornwell's introduction is excellent, and includes a summary of the history of consciousness studies. Margaret Boden organizes a clear discussion of the prospects for a science of consciousness around Jerry Fodor's claim that "nobody has the slightest idea how anything material could be conscious". Searle organizes an equally useful discussion around a series of fallacious arguments against a science of consciousness.

The philosopher of quantum mechanics Jeremy Butterfield provides a rigorous introduction to current philosophical ideas on reduction, supervenience and the definitions of materialism and physicalism. He examines ways in which quantum theory might challenge the physicalist assumptions on which most discussions of the mind/body problem are predicated. Readers unacquainted with the apparatus of contemporary metaphysics, however, may find this paper hard going.

Stephen Rose's lucid discussion of the pitfalls of neurogenetic determinism may seem less relevant to consciousness studies than some of the other papers, but a small slip in Olaf Sporns' introduction to the 'neural Darwinist' approach to the brain shows that Rose's theme is as relevant here as in other areas of psychobiology. Sporns explains that the term 'epigenetic' means "literally 'after' or 'over' the genes". But the use of this term in

Getting to know you

The bonobo (*Pan paniscus*), the fourth living species of great ape along with chimpanzees, gorillas and orangutans, was discovered by science little more than 60 years ago, and our knowledge of this gentle animal, which lives naturally only in remote areas of Zaire, is a fraction of that of the other three species. *Bonobo: The Forgotten Ape* (University of California Press, \$24.95. £16.95) by two Dutchmen, laboratory ethologist Frans de Waal and photographer Frans Lanting, distils the information gathered by the few workers in the field, among them the Japanese primatologists Takayoshi Kano and Suehisa Kuroda (for a review by William McGrew see *Nature* 387, 142–143; 1997).



developmental biology predates the discovery of genetics by some centuries. Epigenetic alludes to *genesis* rather than genes.

Traditionally, epigenesis was the rival hypothesis to preformation — the view that the egg contains a tiny organism that grows in size rather than developing any new structures. The contrasting, epigenetic, view is that simple material components in the egg interact with their environment to produce a complex, functioning organism. Much of today's loose talk about genes as programs veers dangerously close to the preformation theory, and it is this danger that Rose warns us against. If we are still forced to flirt with preformation to get our minds around the fact that growth is a material process, then it is no surprise that we cannot imagine a material process giving rise to consciousness.

The inclusion of two papers by theologians — Nicholas Lash and Fraser Watts — may seem unusual, but their concerns relate quite closely to other papers in the volume. Mary Midgley and William Clocksin both argue that the sciences of the mind must integrate the meaning of human action in its cultural context before they can hope to solve the problem of consciousness. Complementing these views, Watts makes an extended comparison of the relationship between consciousness and materiality with that between theology and science. Lash addresses some of the same difficult questions of

identity using theological ideas developed to deal with the doctrine of the Trinity.

Bernard Balleine and Anthony Dickson's paper introduces another way in which the sciences of the mind might be enriched — by greater attention to the emotions. This idea featured in a recent conference, "Emotion, Consciousness and Qualia", convened by the consciousness studies programme of the University of Arizona at Tucson, whose seminal role Cornwell discusses in his introduction.

I was lucky enough to attend this meeting, where it became clear that many leading neuroscientific investigators of emotion, such as Joseph LeDoux, feel they can answer the important questions about the material basis of emotion without investigating the nature of emotion consciousness. They need not even take a stand on whether the states they investigate, such as states in the rat brain homologous to fear in the human brain, are conscious at all. This rather suggests that Lipton is right and that, if the neuroscience of emotion illuminates the problem of consciousness, it may only be because it helps advance the sciences of the mind, "leaving the identity question to torment philosophers". □

Paul E. Griffiths is in the Unit for History and Philosophy of Science, Faculty of Science, University of Sydney, New South Wales 2006, Australia.

Ornithologist-watching

A Passion for Birds: American Ornithology after Audubon

by Mark V. Barrow Jr

Princeton University Press: 1998. 326 pp.

\$39.50, £24.95

John W. Fitzpatrick

Because they are ubiquitous, variable in shape and behaviour, and easy to study, birds engage our curiosity and provide an extraordinary window onto how nature works. They have been subjects of both human passion and rigorous scientific investigation for centuries, yielding founding principles in ecology, evolutionary biology, population biology, animal behaviour and conservation sciences.

In the United States these fields emerged almost simultaneously around 1900, as the age of exploration and description gave way to more process-oriented inquiry. In Mark Barrow's *A Passion for Birds* we experience this transition in science through the lives and letters of individuals long revered as the fathers of American ornithology (a few mothers also appear, very late). The result is at once copiously footnoted history and lively drama, complete with colourful characters, harrowing escapes, class struggles and personal triumphs.

Barrow exposes a rich portrait of human strength, frailty and ego in the transformation of a popular nineteenth-century hobby into first a scientific discipline and ultimately an influential mainstream profession. Moreover, because of the unique passions that birds elicit in humans, ornithology has long been a field in which the lay citizen plays a large role. Thus, Barrow also exposes a spectacular example of science and popular culture being inextricably — if sometimes unwillingly — bound together. Tensions produced by this nexus provide pertinent reading a century later in light of the burgeoning role of 'citizen-scientists' in ornithology today.

The story opens with vivid accounts of the post-Audubon fascination with bird collecting among pioneer naturalists. The plot is thickest between 1883, when the American Ornithologists' Union (AOU) was founded, and 1930, when an ageing aristocracy of museum-bound ornithologists lost clout as university scientists focused increasingly on the living bird. Barrow concludes with homage to a young immigrant from Germany who struggled during the 1930s to reform the AOU against a senescent and narrow-minded élite. Ernst Mayr, always a champion of the amateur naturalist in ornithology, would become America's most influential professional in the field.

Barrow interweaves several subplots into



Atlantic puffins, from Audubon's *Birds of America*: the starting point for American ornithology.

the emergence of an interdisciplinary science from a field dominated by taxonomists. One explores the profound yet schizoid role played by American ornithologists in the movement to protect nature. It seems curious today that a venerable and staunchly scientific body (the AOU) and an equally venerable body of conservation activists (the Audubon Society) were conceived by the same individuals at exactly the same time. No sooner did the two organizations emerge than they parted company. It took a group of Boston society women outraged by the slaughter of birds for hat plumes to breathe activism into Audubon. In contrast, after early triumphs, the AOU's role in bird conservation became detached and muted, and struggles to define itself even today.

Engagement of the AOU in bird protection waned partly because of public concern over bird collecting, an activity that many scientists viewed as the lifeblood of ornithology. Ironically, this discipline, launched mainly by interested amateurs, then distanced itself from those very amateurs who had begun to *watch* birds, even as the pastime gave rise to new scientific opportunities. Tensions and emotions run high to this day over the issue of scientific collecting. Borne out of the human passions that birds elicit, debate still rages within both professional and amateur ornithology. Modern zealots on both sides should read Barrow's accounts of this century-old argument with care. Unfortunately, his unflattering treatment of scientists who steadfastly defended the importance of collecting voucher specimens reveals an underlying bias that colours many passages in this book.

A related subplot spotlights the rise to power, decades of control and ultimate obsolescence of a small band of men who established and narrowly defined American ornithology for a generation. Museums

provided the first bastion of the profession, but they fostered such a fortress mentality among the practitioners that the very structure of the AOU was built to exclude new blood from a small inner sanctum. Still reflecting a profound desire among professionals to distinguish themselves from growing throngs of amateurs, remnants of this class system are only today breaking down within the AOU.

As an American scientist with — yes — a passion for birds, I was gripped from cover to cover by this book and gained enormous insight into my roots. Scientists and historians of science on both sides of the Atlantic will appreciate this illuminating case history of the human side of a profession's origins. □

John W. Fitzpatrick is in the Cornell Laboratory of Ornithology, Cornell University, Ithaca, New York 14850, USA.

Mediaeval cures and casualties

Medicine in the English Middle Ages

by Faye Getz

Princeton University Press: 1998. 192 pp.

\$32.50, £21.95

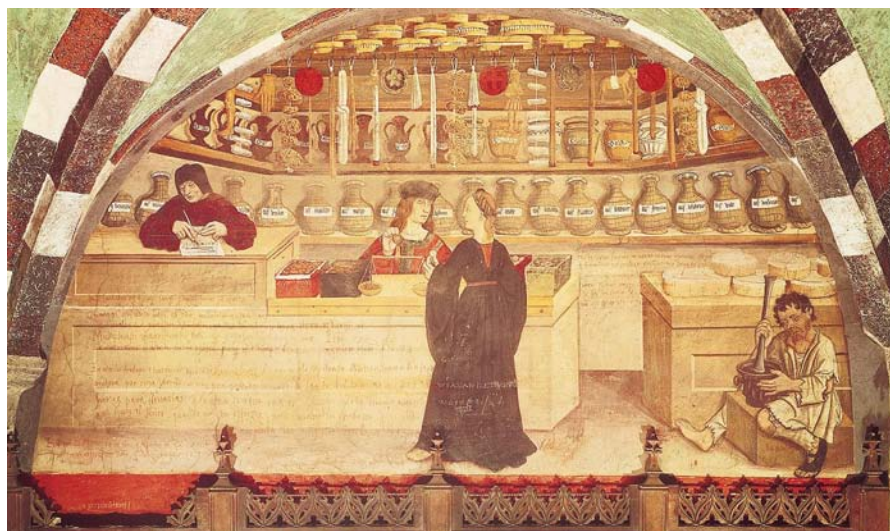
Vivian Nutton

Medicine in England in the Middle Ages offers a melancholy prospect to the historian. There are few great names — John of Gaddesden, Gilbert the Englishman and the surgeon John of Arderne were almost alone in having a reputation on the Continent as well as at home — and the lack of any effective authority, whether from the king, Parliament or the universities, rendered futile all attempts to establish a kingdom-wide organization for either physic or surgery.

Faye Getz takes a more optimistic line in this clearly written and well-organized new survey. She stresses the variety of medical assistance available, from housewives and travelling salesmen up to the tiny number of university-trained doctors serving the very rich as counsellors and cultural assistants, as well as physicians. She raises the question of the effectiveness of the learned élite compared with those who lacked Latin, insisting, perhaps too strongly, that all major treatises in English are only translations from the Latin. We are reminded, too, of how the growing tendency to define proper medicine as university medicine led to the exclusion of those with healing skills but without the qualifications to enter Oxford or Cambridge, especially women and Jews (although the numbers of Jewish healers can hardly be determined).

The English were also stay-at-homes (Wales, Scotland and Ireland are not considered here). Few ventured across the Channel, and fewer still beyond Paris — although more might have been said in the book about the medical learning of such universal scholars and continental travellers as Walter Burley. Foreigners who came to England usually followed a wealthy patron, or confined their attentions to their countrymen resident in London. Some practised astrology, some were clearly in search of quick money, and several are known from their appearances in (not at) court.

Law and medicine are here intertwined. Complaints about incompetence are listed alongside a catalogue of deaths and injuries, some tragic, others simply bizarre, like the case of the collapsing privy. Child abuse is no modern phenomenon, and the elderly and infirm are still seen as a drain



Too far from home: the cures in this Italian pharmacy may well have been unknown in England.

on community resources.

Getz succinctly lists the ways in which those in need of care and assistance might find it in a monastery (although these increasingly excluded the sick) or a hospital (where medical help was rare), or within the extended family. She emphasizes the religious orientation of mediaeval hospitals, which were hospices aiming to secure the long-term salvation of the soul far more than the immediate cure of the body, as well as their sheer variety. Some had a handful of inmates; a very few had more than 100; some excluded almost every kind of patient; others ended up as schools or university colleges.

Less is said about disease and its treatments, and the outbreak of the Black Death is, rightly, played down as the defining point of English mediaeval medicine. Whatever the plague's social consequences, doctors

continued to be consulted both during and after it, and neither in 1348 nor in any of the subsequent outbreaks do surgeons supplant physicians in popular affection, as some have argued.

Getz's book is short (a mere 92 pages of text), abundantly footnoted, at times to excess. It is packed with facts, and its conclusions are well argued. But its brevity is also its weakness. It lacks space to breathe; individual instances pile one on top of the other, where what is needed is a broader exposition. Challenging ideas are raised, but rarely developed, and the wider context is often missing. It is unusual to complain that a book is too short, but this is one instance, proof of the solid quality of the scholarship it contains. □

Vivian Nutton is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

New in paperback

Beyond Calculation: The Next Fifty Years of Computing

edited by Peter J. Denning & Robert M. Metcalfe
Copernicus, £11.50, \$15

"...what sets the book apart from the general run of technology-future books is the authority of its contributors and the tone of restraint that pervades it, when compared with the genre's usual ludicrous extrapolations", Martin Campbell-Kelly, *Nature* 390, 42 (1997).

The Heat Is On: The High Stakes Battle over Earth's Threatened Climate

by Ross Gelbspan
Perseus, \$13

"Ross Gelbspan has produced a marvellously readable but devastatingly candid account of the brutal politics of debunking the scientific method by the opulent vested interests of the fossil-fuel lobby", Tim O'Riordan, *Nature* 389, 685 (1997).

Ptolemy's Almagest

translated and annotated by G. J. Toomer
Princeton University Press, \$39.50

"What Toomer has produced is the best edition in any language, one that will remain the standard preferred text for years to come", Owen Gingerich, *Nature* 308, 789 (1984).

Traces of the Past: Unraveling the Secrets of Archaeology through Chemistry

by Joseph B. Lambert
Perseus, \$18

"Lambert opens a window to a kaleidoscopic view of chemical applications ... The book does not make easy continuous reading, as it is too concentrated and lacks overall themes of interest and excitement. This is a shame, because there are many charming vignettes", Robert Hedges, *Nature* 390, 572 (1997).

Matters of Life and Death: Perspectives on Public Health, Molecular Biology, Cancer and the Prospects for the Human Race

by John Cairns
Princeton University Press, \$14.95, £12.95

"His lucid exposition shows how experiments, observations and calculations support some of the grand conclusions of modern biology. He offers fresh — although sometimes controversial — insights", Joel E. Cohen, *Nature* 387, 565 (1997).

Cartographies of Danger: Mapping Hazards in America

by Mark Monmonier
University of Chicago Press, \$15, £11.95

A distinguished professor of geography at Syracuse University describes how hazard maps can tell us a lot about where to anticipate risks, but also how maps can be dangerously misleading.

Electroluminescence in conjugated polymers

R. H. Friend, R. W. Gymer, A. B. Holmes, J. H. Burroughes, R. N. Marks, C. Taliani, D. D. C. Bradley, D. A. Dos Santos, J. L. Brédas, M. Lögdlund & W. R. Salaneck

Research in the use of organic polymers as the active semiconductors in light-emitting diodes has advanced rapidly, and prototype devices now meet realistic specifications for applications. These achievements have provided insight into many aspects of the background science, from design and synthesis of materials, through materials fabrication issues, to the semiconductor physics of these polymers.

Electroluminescence—the generation of light, other than black-body radiation, by electrical excitation—is a phenomenon that has been seen in a wide range of semiconductors, and for organic semiconductors was first reported for anthracene single crystals in the 1960s^{1,2}. These early studies established that the process responsible for electroluminescence requires injection of electrons from one electrode and holes from the other, the capture of oppositely charged carriers (so-called recombination), and the radiative decay of the excited electron–hole state (exciton) produced by this recombination process.

Development of organic thin-film electroluminescence was spurred on in the 1980s through the work of Tang and Van Slyke³, who demonstrated efficient electroluminescence in two-layer sublimed molecular film devices. These devices consisted of a hole-transporting layer of an aromatic diamine and an emissive layer of 8-hydroxyquinoline aluminium (Alq₃), the structures of which are shown in Fig. 1. Indium–tin oxide, ITO, is used as the hole-injecting electrode and a magnesium–silver alloy as the electron-injecting electrode. A large number of other molecular materials have been used as the charge-transporting or emissive layer in light-emitting diodes (LEDs). There has been a great deal of activity, particularly in Japan, in the development of devices of this general type; very high levels of performance have been reported, with quantum efficiencies (photons out per charge injected) of several per cent⁴.

Since the first report of metallic conductivities in ‘doped’ polyacetylene in 1977⁵, the science of electrically conducting polymers has advanced very rapidly. Applications have taken longer to appear, in spite of some well recognized areas of application for processible conductive materials. More recently, as high-purity polymers have become available, a range of semiconductor devices have been investigated; these include transistors^{6–11}, photodiodes^{12,13} and LEDs^{14–18}. The potential for commercialization is perceived to be high for these semiconductor devices because they are seen to compete in application areas where the market can bear the costs of development. In particular, polymer LEDs now show attractive device characteristics, including efficient light generation, and there are several development programmes now set up to establish procedures for manufacture. The principal interest in the use of polymers lies in the scope for low-cost manufacturing, using solution-processing of film-forming polymers.

In parallel with these development activities, much progress has been made in the understanding of the underlying science that controls the properties of these devices. In comparison with inorganic semiconductors, relatively little is known about the electronic properties of these materials; even the nature of the semiconductor excitations remains controversial. There has been considerable progress made recently in resolving some of those issues which determine the limits to device performance. These relate to the intrinsic electronic structure of the polymers (for example, the energetics of triplet excitons, and the engineering

of high luminescence efficiency in the presence of strong inter-molecular interactions), and also to the scope for ‘optical engineering’ of the devices—to achieve effective coupling between electronic excitations in the polymer and the generated photons. Here we discuss the current understanding of the science and technology, covering materials synthesis, materials properties and device properties. This review is not intended to be comprehensive, and we have (for example) made no attempt to cover the wide range of work of the development of materials with controlled colour. We refer the reader also to recent reviews^{19–22}.

Polymer LED structures

Conjugated polymers derive their semiconducting properties by having delocalized π -electron bonding along the polymer chain. The π (bonding) and π^* (antibonding) orbitals form delocalized valence and conduction wavefunctions, which support mobile charge carriers. The structures of some of those polymers used for LED fabrication is shown in Fig. 2. Electroluminescence from conjugated polymers was first reported in 1990 (ref. 14), using poly(*p*-phenylene vinylene), PPV, as the single semiconductor layer between metallic electrodes, as is illustrated in Fig. 3a. In this structure, the ITO layer functions as a transparent electrode, and allows the light generated within the diode to leave the device. The top electrode is conveniently formed by thermal evaporation of a metal. LED operation is achieved when the diode is biased sufficiently to achieve injection of positive and negative charge

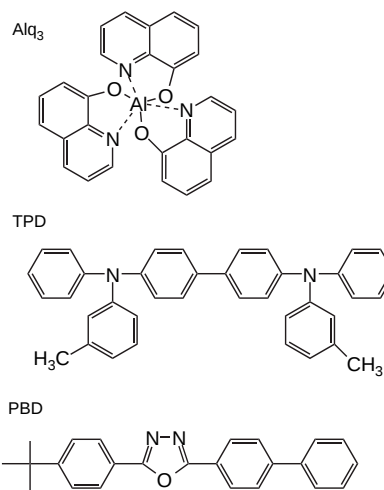


Figure 1 Structures of some molecular semiconductors that have been used in thin-film electroluminescent devices. Alq₃ is used as an electron transport and emissive layer, TPD is used as a hole transport layer, and PBD is used as an electron transport layer.

carriers from opposite electrodes. Capture of oppositely charged carriers within the region of the polymer layer can then result in photon emission.

Diodes of this type can be readily fabricated by solution-processing the semiconducting polymer onto the ITO-coated glass, even though the film thickness is no more than typically 100 nm. Spin-coating from solution has been demonstrated to be capable of producing highly uniform layer thickness, with a thickness variation of no more than a few ångströms spread over several cm². Electrodes are chosen to facilitate charge injection; ITO has a relatively high work function and is therefore suitable for use as a hole-injecting electrode; low-work-function metals such as Al, Mg or Ca are suitable for injection of electrons. The band scheme under forward bias is shown schematically in Fig. 3b.

PPV has an energy gap between π and π^* states of about 2.5 eV, and produces yellow-green luminescence in a band below this energy, with the same spectrum as that produced by photoexcitation (Fig. 4)²³. We note that it shows broadening due to vibronic coupling, as is characteristic for optical transitions in molecular

semiconductors where the excited state is confined to the molecular unit, and is described as an exciton.

The levels of efficiency of the first, simple LEDs based on PPV, which were fabricated with aluminium negative electrodes, were relatively low, of the order of 10⁻⁴ photons generated within the device per electron injected¹⁴ (an internal quantum efficiency of 0.01%). These values have risen rapidly over the past 5 years as improved understanding of the operation of these devices, aided in considerable measure by parallel developments made with sublimed molecular film devices³, has allowed considerable optimization of the device characteristics. The use of negative electrodes with lower work functions was shown to improve efficiency¹⁷, and for single-layer diodes of the type shown in Fig. 3, an external efficiency of about 2 lumens per watt (2 lm W⁻¹) has been reported in devices made with ITO/MEH-PPV/Ca (ref. 24). Other early approaches used to increase efficiency include the use of copolymers based on PPV with higher luminescence efficiencies¹⁵, and the use of heterostructure devices¹⁶. More recently, very much higher efficiencies have been reported for diodes made with similar structures to that shown in Fig. 3, but with a layer of poly(dioxyethylene thienylene) doped with polystyrene sulphonic acid (Fig. 2) between the ITO and the emissive polymer layers. Efficiencies in the green part of the spectrum of up to 16 lm W⁻¹ are reported for diodes using copolymer **4** (Fig. 2) with composition $x = y = 47\%$, $z = 2\%$ (ref. 25), and up to 22 lm W⁻¹ for emissive polymers based on polyfluorene, **7** (Fig. 2)²⁶. Current-density/voltage/luminance characteristics for these polyfluorene-based devices are shown in Fig. 5.

Device operation

A schematic energy-level diagram for a PPV LED under forward bias is shown in Fig. 3b. As described above, polymer LEDs operate by the injection of electrons and holes from negative and positive electrodes, respectively. Electrons and holes capture one another within the polymer film, and form neutral bound excited states (termed excitons). Excitons in conjugated polymers are generally considered to be more strongly localized than excitons in three-dimensional semiconductors, not least because the exciton is substantially confined to a single polymer chain. The spin wavefunction of the exciton, formed from the two spin- $\frac{1}{2}$ electronic charges, can be either singlet ($S = 0$) or triplet ($S = 1$), and a consequence of the confinement of the excitation is that the energy difference between singlet and triplet (the exchange energy) may also be large. Spin-allowed radiative emission (fluorescence) is from the singlet only, and when the exchange energy is large, cross-over from triplet to singlet is unlikely, so that triplet excitons do not produce light emission, other than by indirect processes such as triplet-triplet annihilation, or by phosphorescence.

The internal quantum efficiency η_{int} , defined as the ratio of the number of photons produced within the device to the number of electrons flowing in the external circuit, is given by:

$$\eta_{\text{int}} = \gamma r_{\text{st}} q \quad (1)$$

where γ is the ratio of the number of exciton formation events within the device to the number of electrons flowing in the external circuit, r_{st} is the fraction of excitons which are formed as singlets, and q is the efficiency of radiative decay of these singlet excitons. As we discuss below, the efficiency of radiative decay depends on the device structure, being strongly affected by the photonic structure of the device (for example, the proximity of metallic mirrors).

In order to achieve efficient luminescence, it is therefore necessary to have good balancing of electron and hole currents, efficient capture of electrons and holes within the emissive layer, strong radiative transitions for singlet excitons, and efficient coupling of these excitons to photon states allowed in the device structure. We now discuss these topics.

Electrode interfaces. Injection of charge from most electrode materials requires that charges surmount or tunnel through a

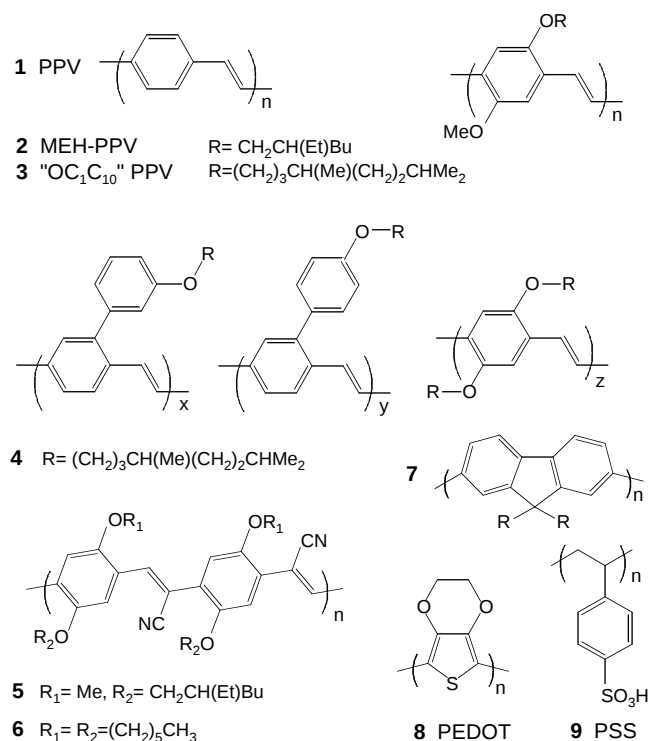


Figure 2 Polymers used in electroluminescent diodes. The prototypical (green) fluorescent polymer is poly(*p*-phenylene vinylene), as shown by **1**. This compound is formed most usually by thermal elimination of tetrahydrothiophene and hydrogen chloride (or bromide for lower temperatures) from a methanol- and water-processible sulphonium precursor⁹¹ prepared by a Wessling condensation polymerization of a bis-sulphonium salt. Condensation polymerization of the bis(halomethyl) monomers affords the two best known (orange-red) solution-processible conjugated polymers MEH-PPV (**2**) and "OC₁C₁₀" PPV (**3**) which have been much used^{20,24}. Copolymers have been widely developed because they allow colour tuning and can show improved luminescence¹⁵; copolymer **4** has recently been reported to show very high electroluminescence efficiency²⁵. Cyano-derivatives of PPV **5** and **6** show increased electron affinities and are used as electron transport materials^{16,56}. Many elegant organometallic coupling procedures have been employed to make poly(phenylene)s and related polymers⁹²⁻⁹⁶. The development of Suzuki condensation coupling, using bis-borates, has provided an important route to high-purity polymers, including poly(dialkylfluorene)s **7** (ref. 97), which show high luminescence efficiencies²⁶. 'Doped' polymers such as poly(dioxyethylene thienylene), PEDOT (**8**), doped with polystyrenesulphonic acid, PSS (**9**), are widely used as hole-injection layers.

barrier at the interface. This is expected on examination of the positions of the electrode metal work functions and the positions of the highest occupied molecular orbitals (HOMO; π orbitals) and the lowest unoccupied molecular orbitals (LUMO; π^* orbitals) in the polymer. For the case of PPV, ITO provides a relatively good match for hole injection, though there is a barrier of around 0.2 eV (ref. 27). However, electron injection is more difficult to achieve without the use of low-work-function, reactive metals such as calcium (for which the barrier, as determined by the difference in work function of the metal and electron affinity of the polymer, is of a similar magnitude).

The nature of interfaces, between the active light-emitting polymer medium and the metal electrode, or between the polymer and the ITO layer, are of paramount importance in determining device performance. The control of these interfaces ultimately may be among the more important determining factors in the eventual success of light-emitting devices²⁸. A combined experiment-theory approach to the study of polymer surfaces and interfaces has been applied to a wide variety of π -conjugated polymers and model molecules, as well as to the early stages of metal-on-polymer (and model molecule) interface formation^{28–30}. During the course of these studies, certain general trends have emerged. One particular issue, which is often neglected, is that there is always 'chemistry' that occurs at the interface²⁸. Even though the metal-on-polymer interface is discussed here, there is a growing body of evidence that additional, equally important, interfacial effects occur at the polymer-on-ITO interface³¹.

The chemistry that occurs at the metal-on-polymer interface varies with the nature of the metal involved, the polymer involved, and especially with the cleanliness of both the materials employed and the vacuum system used in the metallization process. As aluminium and calcium are two of the most commonly used electron-injecting metals, we confine our remarks to effects which

occur during the early stages of interface formation involving these two metals on the surfaces of the PPVs.

The deposition of aluminium atoms upon essentially oxygen-free polymer surfaces leads to cluster formation, and to the formation of covalent bonds at the vinylene-carbon atoms of PPV (and substituted PPVs)³². Although clustering occurs, just as in the case of aluminium atoms on the surfaces of inorganic semiconductors, atomic diffusion takes place into the near-surface region, and covalent bond formation is localized to within a characteristic length scale; this scale is of the order of an electron tunnelling distance, that is 20–30 Å.

In the case of calcium vapour-deposited upon clean surfaces in ultrahigh vacuum (UHV), it was observed clearly, first for a model polyene molecular solid system and subsequently for substituted PPV³³, that calcium atoms diffuse into the near-surface region, donate electrons to the π -system, and form Ca^{2+} ions. The interfacial region between the Ca-metal contact and the polymer has an approximate scale in the range of 20–30 Å (similar to the case of Al atoms). In contrast, when (as a result of exposure of the sample to air) there are large numbers of oxygen-containing species at the surface of, for example, PPV (ref. 34), the reactions are different: first, an interfacial layer of an oxide of calcium is formed when the calcium atoms are deposited in UHV, and then, after the oxygen-containing species have been consumed by these initial calcium atoms, calcium metal is deposited. The scale of this interfacial oxide (insulating) layer also is of the order of 20–30 Å, depending on the details of the surface contamination, chemical impurity of the polymer, and/or the vapour-deposition environment.

Thin barrier layers at the semiconductor/cathode interface can provide improved LED performance, as was shown for molecular film devices, using layers of ionic salts such as LiF evaporated to thicknesses of the order of 1 nm or less³⁵. Such layers probably also occur when cathodes are formed in the presence of a background oxygen pressure. We note that on oxygen-free surfaces of PPV, vapour deposition of calcium in a background pressure of about 10^{-6} mbar of O_2 results in a metallic oxide contact which is observed to increase LED yield, lifetime and luminescence³⁶. The calcium-on-PPV results are summarized in Fig. 6, where either a doped conducting polymer layer is formed on the oxygen-free surface of PPV, or a calcium oxide layer is formed on the surface of PPV containing large amounts of oxygen-containing species.

Turning to the anode: ITO has been the usual choice of material. However, although it has the virtue of optical transparency, it is not a well controlled material. There have been several reports of the use

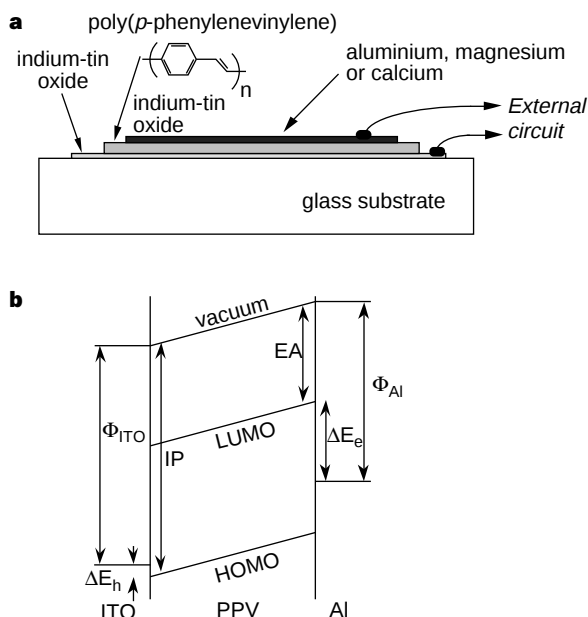


Figure 3 Structure and schematic energy-level diagram of a single-layer polymer electroluminescent diode. **a**, Device structure. **b**, Schematic energy level diagram for an ITO/PPV/Al LED, showing the ionization potential (IP) and electron affinity (EA) of PPV, the work functions of ITO and Al (Φ_{ITO} and Φ_{Al}), and the barriers to injection of electrons and holes (ΔE_e and ΔE_h). There is a small barrier for hole injection from the ITO electrode into the valence band states (or highest occupied molecular orbital, HOMO), and, with aluminium as cathode, a considerably larger barrier for electron injection into the PPV conduction band states (or lowest unoccupied molecular orbital, LUMO).

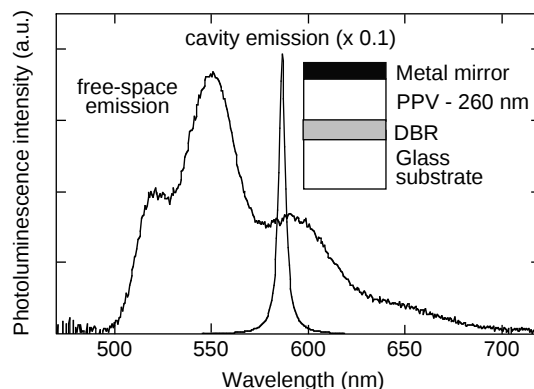


Figure 4 Photoluminescence spectra from PPV. Shown is the optically excited luminescence from PPV as measured for a thin film on a glass substrate (free-space emission) and from a microcavity structure (dimensions as shown) (a.u., arbitrary units). Similar microcavity devices which can be electrically excited are made by forming a layer of ITO onto the distributed Bragg reflector (DBR) mirror before deposition of the polymer layer²³.

of chemically doped conjugated polymers as injection electrodes. Several p-doped conjugated polymers show good environmental stability, including polypyrrole, polythiophene derivatives (Fig. 2) and polyaniline, and such materials have been used as hole-injecting electrodes^{37–39}. Several groups have reported not only improved device efficiency, but also considerably improved device uniformity and longevity^{21,24,40,41}. These doped polymer electrodes have high work functions, thereby providing low barriers for hole injection to the semiconductor layer. It is also likely that there is at least some diffusion of the dopant in the 'electrode' layer to the 'semiconductor' polymer layer, to give a dopant profile into this layer. Though this is clearly desirable in order to achieve easy charge injection, such a diffusion process must be restricted close to the interface, to provide stable operation over long times. We note that most reports of stable operation over long operating times are for devices which use polymeric dopants, which are expected to be relatively immobile. These include polystyrenesulphonic acid, **9**, as used to dope the polythiophene derivative, PEDOT, **8**, shown in Fig. 2.

Pei *et al.*⁴² have shown that with the addition of a salt and ion-transporting material such as poly(ethylene oxide) to the conjugated polymer, very low barriers for charge injection are found. They attribute this to the formation of p-doped regions in the polymer near the positive electrode and an n-doped region near the negative electrode, with the doping arising from electrochemical reactions which occur under drive conditions. Another very important process is the pile-up of mobile ions at both electrode interfaces under the applied field. The very high ionic charge density at the interfaces that can be produced in this way screens the energy barrier for charge injection over very short distances (less than 1 nm), so that tunnelling through this barrier is easy⁴³. The effect of mobile ions on long-term operating stability is not reported, however, although the immobilization of these ions has been discussed⁴⁴.

Charge injection and charge transport. In spite of the very clear evidence for strong chemical interactions between polymer and metal, there is clear evidence that the size of the barriers for electron and hole injection still scales with the electrode work functions. This has been shown in measurements of internal electric field⁴⁵.

The process of charge injection from metal electrodes and the process of charge transport within the polymer layer are difficult to disentangle on the basis of the device electrical characteristics, and the present literature can be confusing. For diodes with large barriers for charge injection (Fig. 3b), the injection of charge, either by thermionic emission or by tunnelling, can certainly limit

current flow^{46,47}. However, charge injection is not considered to limit current flow for LEDs which show good operating characteristics, such as the low turn-on voltage shown in Fig. 5. Instead, current flow is bulk-limited, principally through the build-up of space charge⁴⁸. The space-charge-limited current regime is easily achieved in these structures because the low-field mobilities of charge carriers in relatively disordered molecular semiconductors is very low—less than $10^{-6} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ as modelled by Blom *et al.*⁴⁸.

Modelling of the process of charge injection, taking into account the nature of the available states in the organic semiconductor, produces injection rates consistent with experiment^{49–51}. Modelling of space-charge-limited currents is complicated by the need to take account of a strongly field-dependent mobility⁵², and the influence of bipolar currents on the reduction of the bulk space charge⁵³.

We note that the type of device that produced the characteristics shown in Fig. 5 has both hole-injecting and electron-injecting electrodes with relatively low barriers for charge injection, so that high current densities and concomitant light emission is produced at low voltages. As regards the cathode, calcium is not particularly convenient to handle, but magnesium (alloyed with silver), which is

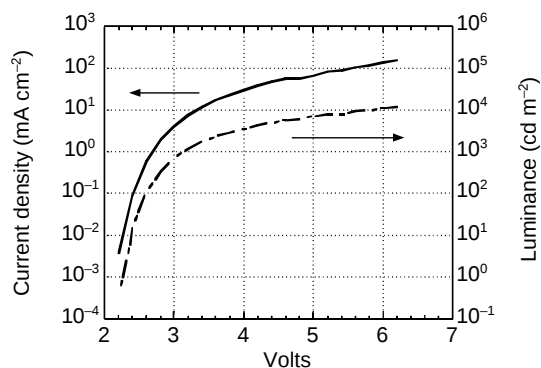


Figure 5 Current density and light output versus drive voltage for green-emitting polymer LEDs based on polyfluorene²⁶. These diodes show turn-on for current and light near 2 V applied bias, and reach 100 cd m^{-2} at 2.6 V, and $1,000 \text{ cd m}^{-2}$ at 3.1 V. Peak efficiencies of 22 lm W^{-1} are achieved near 100 cd m^{-2} . As discussed in the text, current flow is considered to be bulk-limited, rather than injection-limited for devices such as these with low turn-on voltages.

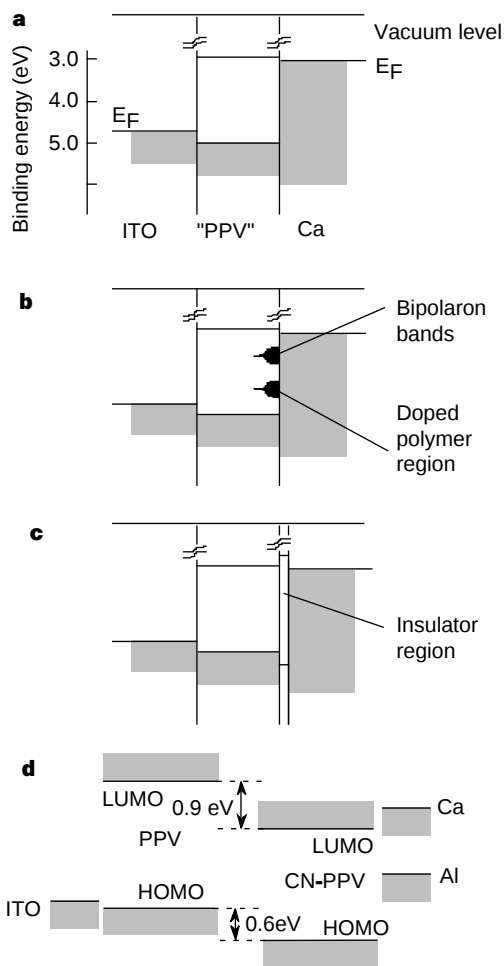


Figure 6 Energy levels for electroluminescent diodes. **a–c**, An ITO-PPV-Ca diode before contact between the three layers, illustrating the energies expected, **a**, from the metal Fermi energies, assuming no chemical interactions at the interface, **b**, after some 'doping' of the interfacial layer of PPV by Ca, setting up 'bipolaron' bands within the PPV semiconductor gap (note that the Fermi energy for the 'doped' PPV lies between the upper bipolaron level and the conduction band), and **c**, after interfacial chemistry which sets up a blocking layer at the interface (as expected in the presence of oxygen). **d**, Energy levels for the components of a two-layer heterojunction diode fabricated with PPV and CN-PPV.

widely used for sublimed molecular film diodes, does not produce such good performance. But aluminium alloyed with low-work-function metals can provide an air-stable material which also gives good performance⁴¹.

Injection and transport of holes from the positive electrode into the bulk of the polymer film must be matched by injection and transport of electrons from the opposite electrode. Recent results from the authors' laboratories obtained with single-layer devices of the type characterized in Fig. 5 indicate that reasonably efficient current balancing can be obtained in these structures. However, for sublimed molecular film devices, the use of two-layer structures to control injection rates of electrons and holes (by introducing barriers for charge transport at the heterojunction between the two semiconductor layers) has been shown to be a very effective means for obtaining current balance^{3,54,55}. Devices of this type made with polymers have been fabricated by a number of groups, most successfully using solution-processible poly(cyanoterephthalylidene)s which are derivatives of PPV with nitrile groups attached to the vinylic carbons (Fig. 2). These complement the existing hole-transporting PPVs. The band scheme for a two-layer device of this type is shown in Fig. 6d. At the interface between the two polymers there are sizeable offsets in the energies of the π orbitals (that is, HOMO) for the PPV and the CN-PPV and also between the π^* orbitals (LUMO) of the PPV and CN-PPV layers. Under forward bias, injection of holes from the ITO into the HOMOs of the PPV layer results in transport of holes to the heterojunction, at which they are confined by the potential barrier which must be surmounted if they are to progress into the CN-PPV layer. Similarly, electrons injected from the negative electrode are confined at the heterojunction by the potential step required to get transfer into the PPV LUMOs. Tunnelling across one or other barrier will allow electron-hole capture and electroluminescence. For these polymers, holes tunnel to the CN-PPV and emission is from this layer (typically in the yellow-red part of the spectrum). External quantum efficiencies as high as 2.5% (for light emitted in the forward direction) are measured with CN-PPV **8** (refs 56, 57).

Electron-hole recombination and light emission. The process of electron-hole capture in these devices is crucial to device operation. In order to get efficient capture in these very thin structures (with a polymer layer total thickness of ~ 100 nm), it is necessary that one or other charge carrier is of very low mobility so that the local charge density is sufficiently high to ensure that the other charge carrier will pass within a collision capture radius of at least one charge. This is certainly enhanced in the heterostructure devices discussed above, where confinement at the heterojunction causes a build up in charge density. Modelling of charge recombination using Langevin theory has been used by several groups^{51,53,58}.

Assuming that the electron-hole capture process is spin-independent (as implicit in the Langevin model), excitons should be formed with spin wavefunctions in the triplet and singlet configurations in the ratio 3:1. For these polymeric semiconductors, there is firm evidence that the triplet exciton is strongly bound with respect to the singlet, so that we expect to lose 75% of the electron-hole pairs to triplet excitons which do not decay radiatively with high efficiency. Experiments to detect the presence of triplet excitons have been performed using measurements of electroluminescence-detected electron-spin resonance⁵⁹ and by measurement of the allowed optical transition between the lowest-lying triplet and a higher-lying triplet at 1.4 eV (ref. 60). This latter work provides an estimate of triplet concentration which is consistent with the 3:1 branching ratio expected within this simple model.

Schemes to make use of the triplet excitons are clearly attractive. One approach is to introduce species that will allow efficient triplet luminescence (phosphorescence). This can be provided by high-atomic-number elements with strong spin-orbit coupling. A platinum-containing porphyrin has been successfully used as a 'dopant' in both a molecular⁶¹ and a polymeric⁶² host; both singlet

and triplet excitons generated in the host material are collected at the porphyrin, which shows efficient phosphorescence. A potential drawback with this approach is that the triplet emission colour is generally considerably red-shifted with respect to the singlet emission of the host, and although this is desirable for red emission^{61,62}, may be problematic for emission colours of green and blue.

Understanding of the nature of the neutral excited states in conjugated polymers has been advanced by a wide range of quantum-chemical calculations^{30,63,64}; inclusion of both electron-electron and electron-phonon terms turns out to be essential to establish a coherent picture of the photoexcitations⁶⁴. Figure 7 illustrates some of the issues concerned. High emission requires that S_1 , the lowest singlet excited state, be strongly coupled (that is, possess a large oscillator strength) with the ground state, S_0 . Soos and co-workers⁶³ have derived the range of parameters related to electron-electron interaction and bond-length alternation in conjugated polymers which ensure that S_1 is of opposite symmetry to S_0 ; this is the case in poly(*p*-phenylene) and poly(*p*-phenylene vinylene) but not in polyacetylene. The last is therefore not luminescent.

The nature of the S_1 state itself is of importance, and has long been debated. A broad consensus has now emerged: S_1 is viewed as a polaron-exciton of weak to intermediate binding (at most a few tenths of an electron volt)^{65,66}. (By 'polaron-exciton', we mean an exciton associated with a local geometry (lattice) relaxation, a feature fully confirmed by the strong vibronic progression seen in optical absorption and, as seen in Fig. 4, emission.)

Radiative emission from the singlet exciton is conveniently measured as photoluminescence (this depends on only the factor q in equation (1)). There has been considerable progress in the search for conjugated polymers which show efficient photoluminescence in the solid state. Measurements obtained using an integrating sphere have been reported⁶⁷. Luminescence efficiency in the solid state tends to be lower than that measured for isolated molecules; this is due both to exciton migration to quenching sites, and to interchain interactions which produce lower-energy excited states which are not strongly radiatively coupled to the ground state⁶⁸. An important source of quenching sites is provided by chemical doping, and many of the first conjugated polymers to become available were sufficiently doped that luminescence yields were very low. However, the PPVs prepared by the precursor routes show photoluminescence efficiencies between 27% (ref. 67) and

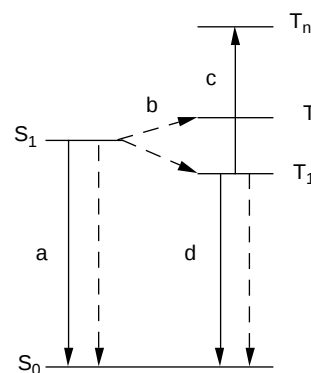


Figure 7 Schematic representation of pathways for singlet decay as well as triplet excitation and decay. Solid arrows represent radiative processes, corresponding to absorption or emission of light; dashed lines denote non-radiative processes. Fluorescence (a), intersystem crossing (b), photoinduced triplet-triplet absorption (c), and phosphorescence (d) are represented. S_0 is the ground (singlet) state; S_1 is the first excited singlet; T_1 the first excited triplet; and T_i and T_n are higher-lying triplet states.

80% (ref. 41), and some of the soluble derivatives, such as the cyano-substituted polymers (Fig. 2) and copolymers, show efficiencies of around 50%.

There is considerable evidence that interchain interactions can significantly modify the energetics of exciton formation in conjugated polymers. For example, interchain interactions are present in the ground state of some of the CN-PPV (ref. 69). The situation for PPV is controversial at present. Measurements at Cambridge of the photoluminescence efficiency (27%) and the decay rate (0.32 ns) are consistent with initial photogeneration of singlet, intrachain excitons which then decay both radiatively (with a lifetime of the order of 1 ns) and non-radiatively⁶⁷. However, Rothberg and co-workers have presented evidence for the formation of interchain charge-separated states⁷⁰, though this might be associated with partially oxidized polymer, for which carbonyl groups act as electron traps and facilitate charge separation, as suggested by measurements of photoluminescence excitation spectra⁷¹. There are several studies of the effects of interchain order and interactions in well-controlled model oligomer systems^{72,73}.

Photonic structure effects. It is now recognized that the coupling of electronic excitations to photon states is strongly affected by the physical structure of the diode. In device structures of the type discussed here, the presence of the metallic cathode provides a mirror which modifies the pattern of the electromagnetic modes near the cathode, setting up standing-wave states. This has the effect of reducing the radiative emission rates for polymer chains which are placed at the nodal spacings from the cathode, and in addition, there can be energy transfer into plasmon modes in the metal⁷⁴. Studies of model structures, in which a thin semiconducting polymer layer is spaced away from a metallic layer by a dielectric layer (silicon oxide), show the effects of plasmon losses at short distances (<30 nm); strong oscillations due to the standing wave states are also found, with an optimum spacing near 60 nm (dependent on the emission wavelength and the refractive index of the spacer layer)⁵⁷. This is an important consideration for the design of efficient diodes; in this respect, heterojunction devices of the type developed by Tang³ are particularly convenient, as the recombination process is close to the heterojunction, and this can be placed at the correct position with respect to the cathode electrode by selection of the thickness of the electron-transport layer. For diodes of the type characterized in Fig. 5, where electron-hole recombination takes place over a broader region extending from the cathode, some loss of performance is inevitable. Experimentally, we find an optimum thickness for the polymer layer of around 70 nm.

The polymer LED structure is easily adapted to function as a microcavity device by the addition of a second mirror. This type of structure has been investigated with many materials systems, including sublimed molecular films^{75,76} and conjugated polymers^{23,77,78}. Both metallic and dielectric (distributed Bragg reflector, DBR) mirrors have been used, as is illustrated by the structure shown in Fig. 4. The Fabry-Pérot etalon formed by the two mirrors defines the allowed cavity electromagnetic modes, which have narrow linewidths for cavities formed with high-reflectivity mirrors. Emission from the excited polymer is only possible into these modes, and only then when the polymer chromophore is placed away from nodes in the standing-wave electromagnetic field pattern. The spectrum shown in Fig. 4 shows the very narrow emission that can be readily achieved. In addition, there is an undesirable angular dispersion of the wavelength (with a blue-shift off-axis), and a considerable redistribution of the angular dependence of the intensity. The latter may be useful when emission is required only in the forward direction; very considerable intensity enhancements in this direction, as seen in Fig. 4 for example, have been reported^{23,76}. Such microcavity structures are also capable of supporting optically pumped lasing⁷⁹. We note that considerable efforts are being made to investigate whether it is feasible to

construct a polymer injection laser. In spite of the demonstration of high peak current densities (above 1,000 A cm⁻²), there are considerable intrinsic problems due excited state absorption from injected carriers^{80,81}.

Efficiency and stability

The model for device operation of polymer LEDs sets some limits to the efficiency of operation. From equation (1), we see that three factors are involved. (1) Charge balancing, covered by γ , is in principle capable of reaching high values, particularly for heterojunction devices. (2) Recombination, covered by r_{st} , is limited to 25% if the creation of excitons is in the ratio 3:1 triplet:singlet. At present, there is only limited information on the energetics of triplet excitons in both polymer and sublimed film devices; though calculations suggest an exchange energy of several tenths of an electron volt for PPV⁸², experimental evidence has been hard to find. There may be considerable scope for design of materials with low exchange energies which would show intersystem crossing back to the singlet and therefore more efficient generation of emissive excitons, and as mentioned above, direct use of triplets in phosphorescent emitters^{61,62}. (3) Radiative emission from excitons, covered by $r_{st}q$, has now been measured in realistic conditions (the presence of electrodes), and there is rapid progress in obtaining high efficiencies. Values well in excess of 50% are realistic.

The values of efficiency reported for polymer devices, now above 20 lm W⁻¹ for green emission (Fig. 5) are now very high, and compare favourably to those reported for sublimed molecular film devices. Among these, devices made with quinacridine-doped Alq₃ as the (green) emissive layer^{55,83} are reported to show luminous efficiencies of 10 lm W⁻¹ at a luminance near 100 candelas per metre² (100 cd m⁻²), and devices with bluer emission show 6 lm W⁻¹ (refs 84, 85). On the basis that efficiency is limited to 25% by the singlet-triplet ratio, these numbers are very high, and represent values for the product γq in excess of 50%.

Probably the most critical performance characteristic for organic LEDs is that of device lifetime, both storage and operational. Operating lifetimes in excess of 3,000 hours, and, generally, in excess of 10,000 hours, are required for most applications, and storage times of 5 years or more are required. Important progress has been reported in this area. First, it is clear that all materials, both polymeric and molecular, will photooxidize⁸⁶ in the presence of water and/or oxygen (the former is now found to play a very important role⁸⁷). Devices must therefore be encapsulated against ingress of water and oxygen, and though this has been achieved for devices made on glass, there are unsolved problems in achieving adequate performance from flexible substrates¹⁸ (important primarily to allow reel-to-reel coating).

There are many studies of degradation processes now in the literature²¹. Devices under operation often show deterioration through formation of 'black spots', which are commonly attributed to the presence of pinholes in the cathode metal, and to ingress of water/oxygen which can allow delamination of the cathode from the semiconductor^{21,88}. It has also been suggested that the oxygen present in ITO may be involved⁴⁰.

Limited results on measured lifetime are available in the literature. The best results reported for sublimed molecular films, based on the Kodak structure, give more than 10,000 hours to half initial brightness (300 cd m⁻²) for devices made with quinacridine-doped Alq₃ as the emissive layer, and an Al/Li alloy as cathode⁸³. For the polymer devices, there are detailed studies of devices made with "OC₁C₁₀" PPV reported by the Philips group, under ambient and also accelerated testing conditions²⁴. Operation at display-level brightnesses to beyond 10,000 hours is routinely achieved with little loss of brightness under constant current drive. The most problematic ageing process is the gradual rise in drive voltage required to maintain constant current; this can be²⁴ as much as 70% of the initial drive voltage over a period of 5,000 hours. This is

considered to be due to a reduced carrier mobility in the polymer, and has been associated with irreversible chemical changes; whether these are intrinsic or due to extrinsic factors, caused for example by poor encapsulation, is not reported at present.

Towards applications

The performance of organic LEDs meets many of the targets necessary for applications in displays. Backlighting and segmented displays have been identified for early products by Philips, and this company has announced the setting up of a pilot line at Heerlen. Passive matrix-addressed displays are attractive, as the device construction is relatively simple. However, this does require that pixels are driven, row by row, under pulsed conditions, and both resistive losses in the conductive tracks and efficiency reductions for the diodes at high current densities limit the pixel number. Such displays have been developed with the sublimed molecular film devices as alphanumeric displays, (64×256 ; ref. 83), where average luminances of around 100 cd m^{-2} are required (this is a typical computer monitor level), and as far as a 1/4-sized VGA display (320×240 ; refs 84, 85). Extension to a full-colour graphic display (for computer monitors and for video display) is very attractive, not least because organic LEDs provide full viewing angle and video-rate response times (in contrast to current liquid-crystal display technology). However, this required red, green and blue colours with appropriate chromaticity, methods for colour patterning, and also new addressing schemes. Rapid progress is being made with these three problems. Development of full colour has been reported, both for sublimed films^{84,85} and for polymers²⁶. Solution-processing of polymers offers new methods for colour patterning, among which there is particular interest in ink-jet printing, to place separated pixels of red-, green- and blue-emitting polymers onto the prepared substrate. This is being developed by Seiko-Epson and Cambridge Display Technology²⁶, and the use of ink-jet printing has also been reported by other groups^{89,90}. Active-matrix transistor arrays, modified from those at present used for liquid-crystal displays, can now provide sufficient current-driving capability to meet the requirements of organic LEDs. This is made possible by the improved properties of polycrystalline silicon compared to those of amorphous silicon, and demonstrator active-matrix polymer displays have been made²⁶. □

R.H.F. and R.W.G. are at the Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK. A.B.H. is at the Melville Laboratory for Polymer Synthesis, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK. R.H.F. and J.H.B. are at Cambridge Display Technology, 181A Huntingdon Road, Cambridge CB3 0DJ, UK. R.N.M. and C.T. are at the Istituto di Spettroscopia Molecolare, CNR, Via Gobetti, 101, 40129 Bologna, Italy. D.D.C.B. is at the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK. D.A.D.S. and J.L.B. are at the Centre de Recherche en Electronique et Photonique Moléculaires, Université de Mons-Hainaut, Place du Parc, 20, B-7000 Mons, Belgium. M.L. and W.R.S. are at the Department of Physics, Linköping University, S-58183 Linköping, Sweden.

- Pope, M., Kallmann, H. & Magnante, P. Electroluminescence in organic crystals. *J. Chem. Phys.* **38**, 2042–2043 (1963).
- Helfrich, W. & Schneider, W. G. Recombination radiation in anthracene crystals. *Phys. Rev. Lett.* **14**, 229–231 (1965).
- Tang, C. W. & Van Slyke, S. A. Organic electroluminescent diodes. *Appl. Phys. Lett.* **51**, 913–915 (1987).
- Adachi, C., Tsutsui, T. & Saito, S. Organic electroluminescent device having a hole conductor as an emitting layer. *Appl. Phys. Lett.* **55**, 1489–1491 (1989).
- Chiang, C. K. et al. Electrical conductivity in doped polyacetylene. *Phys. Rev. Lett.* **39**, 1098–1101 (1977).
- Garnier, F., Hajlaoui, R., Yassar, A. & Srivastava, P. All-polymer field-effect transistor realized by printing techniques. *Science* **265**, 1684–1686 (1994).
- Ostoj, P. et al. Electrical characteristics of field-effect transistors formed with ordered alpha-sexithienyl. *Synth. Met.* **54**, 447–452 (1993).
- Torsi, L., Dodabalapur, A., Rothberg, L. J., Fung, A. W. P. & Katz, H. E. Intrinsic transport-properties and performance limits of organic field-effect transistors. *Science* **272**, 1462–1464 (1996).
- Yang, Y. & Heeger, A. J. A new architecture for polymer transistors. *Nature* **372**, 344–346 (1994).
- Brown, A. R., Pomp, A., Hart, C. M. & Deleuw, D. M. Logic gates made from polymer transistors and their use in ring oscillators. *Science* **270**, 972–974 (1995).
- Sirringhaus, H., Tessler, N. & Friend, R. H. Integrated optoelectronic circuits based on conjugated polymers. *Science* **280**, 1741–1744 (1998).
- Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells—enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science* **270**, 1789–1791 (1995).
- Granström, M. et al. Laminated fabrication of polymeric photovoltaic diodes. *Nature* **395**, 257–260 (1998).
- Burroughes, J. H. et al. Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
- Burn, P. L. et al. Chemical tuning of electroluminescent copolymers to improve emission efficiencies and allow patterning. *Nature* **356**, 47–49 (1992).
- Greenham, N. C., Moratti, S. C., Bradley, D. D. C., Friend, R. H. & Holmes, A. B. Efficient polymer-based light-emitting diodes based on polymers with high electron affinities. *Nature* **365**, 628–630 (1993).
- Braun, D. & Heeger, A. J. Visible light emission from semiconducting polymer diodes. *Appl. Phys. Lett.* **58**, 1982–1984 (1991).
- Gustafsson, G. et al. Flexible light-emitting diodes made from soluble conducting polymers. *Nature* **357**, 477–479 (1992).
- Greenham, N. C. & Friend, R. H. in *Solid State Physics* (eds Ehrenreich, H. & Spaepen, F. A.) 2–150 (Academic, San Diego, USA, 1995).
- Salbeck, J. Electroluminescence with Organic Compounds. *Ber. Bunsenges. Phys. Chem.* **100**, 1667–1677 (1996).
- Sheats, J. R. et al. Organic electroluminescent devices. *Science* **273**, 884–888 (1996).
- Kraft, A., Grimsdale, A. C. & Holmes, A. B. Electroluminescent conjugated polymers—seeing polymers in a new light. *Angew. Chem. Int. Ed. Engl.* **37**, 402–428 (1998).
- Grüner, J., Cacialli, F. & Friend, R. H. Emission enhancement in single-layered conjugated polymer microcavities. *J. Appl. Phys.* **80**, 207–215 (1996).
- Liedenbaum, C., Croonen, Y., van de Weijer, P., Vleggaar, J. & Schoo, H. Low voltage operation of large area polymer LEDs. *Synth. Met.* **91**, 109–111 (1997).
- Spreitzer, H. et al. Soluble phenyl-substituted PPVs—New materials for highly efficient polymer LEDs. *Adv. Mater.* **10**, 1340–1344 (1998).
- Lacey, D. High-efficiency polymer light-emitting diodes. 9th International Workshop on Inorganic and Organic Electroluminescence, Oregon, Bend, USA, September 13–17, 1998. See also (<http://www.cdtltd.co.uk>)
- Brown, A. R. et al. Poly(p-phenylene vinylene) light-emitting diodes: enhanced electroluminescence efficiency through charge carrier confinement. *Appl. Phys. Lett.* **61**, 2793–2795 (1992).
- Salaneck, W. R. & Brédas, J. L. The metal-on-polymer interface in polymer light-emitting-diodes. *Adv. Mater.* **8**, 48–52 (1996).
- Salaneck, W. R., Stafström, S. & Brédas, J. L. *Conjugated Polymer Surfaces and Interfaces* (Cambridge Univ. Press, 1996).
- Brédas, J. L. Conjugated polymers and oligomers—designing novel materials using a quantum-chemical approach. *Adv. Mater.* **7**, 263–274 (1995).
- Johansson, N. et al. A study of the ITO-on-PPV interface using photoelectron spectroscopy. *Synth. Met.* **92**, 207–211 (1998).
- Logdlund, M. & Brédas, J. L. Theoretical-studies of the interaction between aluminum and poly(p-phenylenevinylene) and derivatives. *J. Chem. Phys.* **101**, 4357–4364 (1994).
- Dannetun, P. et al. Interface formation between poly(2,5-dihexyl-p-phenylenevinylene) and calcium—implications for light-emitting diodes. *Synth. Met.* **67**, 133–136 (1994).
- Gao, Y., Park, K. T. & Hsieh, B. R. Interface formation of Ca with poly(p-phenylene vinylene). *J. Appl. Phys.* **73**, 7894–7899 (1993).
- Hung, L. S., Tang, C. W. & Mason, M. G. Enhanced electron injection in organic electroluminescence devices using an Al/LiF electrode. *Appl. Phys. Lett.* **70**, 152–154 (1997).
- Bröms, P., Birgersson, J., Johansson, N., Logdlund, M. & Salaneck, W. R. Calcium electrodes in polymer LEDs. *Synth. Met.* **74**, 179–181 (1995).
- Hayashi, S., Etoh, H. & Saito, S. Electroluminescence of perylene films with a conducting polymer as an anode. *Jpn J. Appl. Phys.* **25**, 773–775 (1986).
- Nakano, T., Doi, S., Noguchi, T., Ohnishi, T. & Iyechika, Y. Organic electroluminescent device. US Patent 5,317,169 (1994).
- Yang, Y. & Heeger, A. J. Polyaniline as a transparent electrode for polymer light-emitting diodes: lower operating voltage and higher efficiency. *Appl. Phys. Lett.* **64**, 1245–1247 (1994).
- Karg, S., Scott, J. C., Salem, J. R. & Angelopoulos, M. Increased brightness and lifetime of polymer light-emitting-diodes with polyaniline anodes. *Synth. Met.* **80**, 111–117 (1996).
- Carter, J. C. et al. Operating stability of light-emitting polymer diodes based on poly(p-phenylene vinylene). *Appl. Phys. Lett.* **71**, 34–36 (1997).
- Pei, Q. B., Yu, G., Zhang, C., Yang, Y. & Heeger, A. J. Polymer light-emitting electrochemical-cells. *Science* **269**, 1086–1088 (1995).
- de Mello, J. C., Tessler, N., Graham, S. C. & Friend, R. H. Ionic space-charge effects in polymer light-emitting diodes. *Phys. Rev. B* **57**, 12951–12963 (1998).
- Gao, J., Yu, G. & Heeger, A. J. Polymer light-emitting electrochemical cells with frozen p-i-n junction. *Appl. Phys. Lett.* **71**, 1293–1295 (1997).
- Campbell, I. H., Hagler, T. W., Smith, D. L. & Ferraris, J. P. Direct measurement of conjugated polymer electronic excitation-energies using metal/polymer/metal structures. *Phys. Rev. Lett.* **76**, 1900–1903 (1996).
- Marks, R. N., Bradley, D. D. C., Jackson, R. W., Burn, P. L. & Holmes, A. B. Charge injection and transport in poly(p-phenylene vinylene) light-emitting-diodes. *Synth. Met.* **57**, 4128–4133 (1993).
- Parker, I. D. Carrier tunneling and device characteristics in polymer light-emitting-diodes. *J. Appl. Phys.* **75**, 1656–1666 (1994).
- Blom, P. W. M., De Jong, M. J. M. & Vleggaar, J. J. M. Electron and hole transport in poly(p-phenylene vinylene) devices. *Appl. Phys. Lett.* **68**, 3308–3310 (1996).
- Davids, P. S., Kogan, S. M., Parker, I. D. & Smith, D. L. Charge injection in organic light-emitting-diodes—tunneling into low mobility materials. *Appl. Phys. Lett.* **69**, 2270–2272 (1996).
- Conwell, E. M. & Wu, M. W. Contact injection into polymer light-emitting diodes. *Appl. Phys. Lett.* **70**, 1867–1869 (1997).
- Bassler, H. Injection, transport and recombination of charge carriers in organic light-emitting diodes. *Polym. Adv. Technol.* **9**, 402–418 (1998).
- Blom, P. W. M., de Long, M. J. & van Munster, M. G. Electric-field and temperature dependence of the hole mobility in poly(p-phenylene vinylene). *Phys. Rev. B* **55**, 1–3 (1997).
- Blom, P. W. M., de Jong, M. J. M. & Breedijk, S. Temperature dependent electron-hole recombination in polymer light-emitting diodes. *Appl. Phys. Lett.* **71**, 930–932 (1997).
- Tang, C. W., Van Slyke, S. A. & Chen, C. H. Electroluminescence of doped organic thin films. *J. Appl. Phys.* **65**, 3610–3616 (1989).
- Shi, J. M. & Tang, C. W. Doped organic electroluminescent devices with improved stability. *Appl. Phys. Lett.* **70**, 1665–1667 (1997).
- Baigent, D. R. et al. Light-emitting diodes fabricated with conjugated polymers—recent progress. *Synth. Met.* **67**, 3–10 (1994).
- Becker, H., Burns, S. E. & Friend, R. H. The effect of metal films on the photoluminescence and electroluminescence of conjugated polymers. *Phys. Rev. B* **56**, 1893–1905 (1997).
- Scott, J. C., Karg, S. & Carter, S. A. Bipolar charge and current distributions in organic light-emitting diodes. *J. Appl. Phys.* **82**, 1454–1460 (1997).

59. Swanson, L. S. *et al.* Electroluminescence-detected magnetic-resonance study of poly(paraphenylene-vinylene) (PPV)-based light-emitting diodes. *Phys. Rev. B* **46**, 15072–15077 (1992).
60. Brown, A. R. *et al.* Optical spectroscopy of triplet excitons and charged excitations in poly(p-phenylenevinylene) light-emitting diodes. *Chem. Phys. Lett.* **210**, 61–66 (1993).
61. Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**, 151–154 (1998).
62. Cleave, V. M., Yahioglu, G., Le Barny, P., Friend, R. H. & Tessler, N. Harvesting of singlet and triplet energy in polymer LEDs. *Adv. Mater.* (in the press).
63. Soos, Z. G., Galvao, D. S. & Etemad, S. Fluorescence and excited-state structure of conjugated polymers. *Adv. Mater.* **6**, 280–287 (1994).
64. Chandross, M. *et al.* Excitons in poly(para-phenylenevinylene). *Phys. Rev. B* **50**, 14702–14705 (1994).
65. Brédas, J. L., Cornil, J. & Heeger, A. J. The exciton binding-energy in luminescent conjugated polymers. *Adv. Mater.* **8**, 447–452 (1996).
66. Alvarado, S. F., Seidler, P. F., Lidzey, D. G. & Bradley, D. D. C. Direct determination of the exciton binding energy of conjugated polymers using a scanning tunneling microscope. *Phys. Rev. Lett.* **81**, 1082–1085 (1998).
67. Greenham, N. C. *et al.* Measurement of absolute photoluminescence quantum efficiencies in conjugated polymers. *Chem. Phys. Lett.* **241**, 89–96 (1995).
68. Cornil, J., dos Santos, D. A., Crispin, X., Silbey, R. & Brédas, J. L. Influence of interchain interactions on the absorption and luminescence of conjugated oligomers and polymers: a quantum-chemical characterization. *J. Am. Chem. Soc.* **120**, 1289–1299 (1998).
69. Samuel, I. D. W., Rumbles, G. & Collison, C. J. Efficient interchain photoluminescence in a high-electron-affinity conjugated polymer. *Phys. Rev. B* **52**, 11573–11576 (1995).
70. Rothberg, L. J. *et al.* Photophysics of phenylenevinylene polymers. *Synth. Met.* **80**, 41–58 (1996).
71. Harrison, N. T., Hayes, G. R., Phillips, R. T. & Friend, R. H. Singlet intrachain exciton generation and decay in poly(p-phenylene vinylene). *Phys. Rev. Lett.* **77**, 1881–1884 (1996).
72. Muccini, M. *et al.* Polarized fluorescence in alpha-sexithienyl single crystal at 4.2 K. *J. Chem. Phys.* **108**, 7327–7333 (1998).
73. Gill, R. E., Meetsma, A. & Hadzioannou, G. Two novel thermotropic liquid-crystalline substituted oligo(p-phenylene-vinylene)s—single-crystal x-ray determination of an all-trans oligomeric PPV. *Adv. Mater.* **8**, 212–218 (1996).
74. Chance, R. R., Prock, A. & Silbey, R. in *Advances in Chemical Physics* (eds Prigogine, I. & Rice, S. A.) 1–65 (Wiley, New York, 1978).
75. Tsutsui, T., Takada, N., Saito, S. & Ogino, E. Sharply directed emission in organic electroluminescent diodes with an optical-microcavity structure. *Appl. Phys. Lett.* **65**, 1868–1870 (1994).
76. Jordan, R. H., Rothberg, L. J., Dodabalapur, A. & Slusher, R. E. Efficiency enhancement of microcavity organic light-emitting-diodes. *Appl. Phys. Lett.* **69**, 1997–1999 (1996).
77. Fisher, T. A. *et al.* Electroluminescence from a conjugated polymer microcavity structure. *Appl. Phys. Lett.* **67**, 1355–1357 (1995).
78. Lemmer, U. *et al.* Microcavity effects in a spin-coated polymer 2-layer system. *Appl. Phys. Lett.* **66**, 1301–1303 (1995).
79. Tessler, N., Denton, G. J. & Friend, R. H. Lasing in conjugated polymer microcavities. *Nature* **382**, 695–697 (1996).
80. Tessler, N., Harrison, N. T. & Friend, R. H. High peak brightness polymer light-emitting diodes. *Adv. Mater.* **10**, 64–68 (1998).
81. Lidzey, D. G., Bradley, D. D. C., Alvarado, S. F. & Seidler, P. F. Electroluminescence in polymer films. *Nature* **386**, 135 (1997).
82. Beljonne, D., Shuai, Z., Friend, R. H. & Brédas, J. L. Theoretical investigation of the lowest singlet and triplet states in poly(para phenylene vinylene) oligomers. *J. Chem. Phys.* **102**, 2042–2049 (1995).
83. Wakimoto, T. *et al.* Stability characteristics of quinaclidone and coumarin molecules as guest dopants in the organic LEDs. *Synth. Met.* **91**, 15–19 (1997).
84. Hosokawa, C. *et al.* Organic multi-color electroluminescence display with fine pixels. *Synth. Met.* **91**, 3–7 (1997).
85. Miyaguchi, S. *et al.* OLED full-color passive matrix display. 9th International Workshop on Inorganic and Organic Electroluminescence, Oregon, Bend, USA, September 13–17, 1998.
86. Scurlock, R. D., Wang, B. J., Ogilby, P. R., Sheats, J. R. & Clough, R. L. Singlet oxygen as a reactive intermediate in the photodegradation of an electroluminescent polymer. *J. Am. Chem. Soc.* **117**, 10194–10202 (1995).
87. Xing, K. Z. *et al.* The interaction of poly(p-phenylene-vinylene) with air. *Adv. Mater.* **8**, 971–974 (1996).
88. Burrows, P. E. *et al.* Reliability and degradation of organic light-emitting devices. *Appl. Phys. Lett.* **65**, 2922–2944 (1994).
89. Hebner, T. R., Wu, C. C., Marcy, D., Lu, M. H. & Sturm, J. C. Ink-jet printing of doped polymers for organic light emitting devices. *Appl. Phys. Lett.* **72**, 519–521 (1998).
90. Bharathan, J. & Yang, Y. Polymer electroluminescent devices processed by inkjet printing: I. Polymer light-emitting logo. *Appl. Phys. Lett.* **72**, 2660–2662 (1998).
91. Burn, P. L. *et al.* Precursor route chemistry and electronic properties of poly(1,4-phenylenevinylene), poly(2,5-dimethyl-1,4-phenylenevinylene), and poly(2,5-dimethoxy-1,4-phenylenevinylene). *J. Chem. Soc. Perkin Trans 1* 3225–3231 (1992).
92. Remmers, M. *et al.* The optical, electronic, and electroluminescent properties of novel poly(p-phenylene)-related polymers. *Macromolecules* **29**, 7432–7445 (1996).
93. Scherf, U. & Müllen, K. A soluble ladder polymer via bridging of functionalised poly(p-phenylene)-precursors. *Makromol. Chem. Rapid Commun.* **12**, 489–797 (1991).
94. Yang, Y. *et al.* Efficient blue polymer light-emitting-diodes from a series of soluble poly(paraphenylene)s. *J. Appl. Phys.* **79**, 934–939 (1996).
95. Birgersson, J. *et al.* Efficient blue-light emitting devices from conjugated polymer blends. *Adv. Mater.* **8**, 982–985 (1996).
96. Klarner, G., Davey, M. H., Chen, W. D., Scott, J. C. & Miller, R. D. Colorfast blue-light-emitting random copolymers derived from di-n-hexylfluorene and anthracene. *Adv. Mater.* **10**, 993–999 (1998).
97. Janietz, S. *et al.* Electrochemical determination of the ionization potential and electron affinity of poly(9,9-dioctylfluorene). *Appl. Phys. Lett.* **73**, 2453–2455 (1998).

Acknowledgements. We thank the European Commission (ESPRIT programme 8013 LEDFOS) for support. Research in Mons is partly supported by the Belgian Federal Government (Pôle d'Attraction Interuniversitaire en Chimie Supramoléculaire et Catalyse), FRNS-FRFC, and an IBM Academic Joint Study.

Correspondence should be addressed to R.H.F. at the University of Cambridge (e-mail: rhf10@cam.ac.uk).

A single myosin head moves along an actin filament with regular steps of 5.3 nanometres

Kazuo Kitamura^{*†‡}, Makio Tokunaga^{*§}, Atsuko Hikikoshi Iwane^{†‡} & Toshio Yanagida^{*†‡||}

^{*}Yanagida BioMotron Project, ERATO, JST, 2-4-14 Senba-Higashi, Mino, Osaka 562-0035, Japan

[†]Department of Biophysical Engineering, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

[‡]Department of Physiology I, Osaka University Medical School, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

[§]Structural Biology Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

^{||}Single Molecule Processes Project, ICORP, JST, 2-4-14 Senba-Higashi, Mino, Osaka 562-0035, Japan

Actomyosin, a complex of actin filaments and myosin motor proteins, is responsible for force generation during muscle contraction. To resolve the individual mechanical events of force generation by actomyosin, we have developed a new instrument with which we can capture and directly manipulate individual myosin subfragment-1 molecules using a scanning probe. Single subfragment-1 molecules can be visualized by using a fluorescent label. The data that we obtain using this technique are consistent with myosin moving along an actin filament with single mechanical steps of approximately 5.3 nanometres; groups of two to five rapid steps in succession often produce displacements of 11 to 30 nanometres. This multiple stepping is produced by a single myosin head during just one biochemical cycle of ATP hydrolysis.

Studies of the actomyosin motor have entered a new phase in the past few years. Structures of the actin monomer¹ and the myosin head², determined by X-ray crystallography, have provided a framework for understanding the interaction between the myosin head and the actin filament. The 'tilting crossbridge' model of force generation^{3,4} has been refined, on the basis of the myosin crystal structure, into the 'lever-arm' hypothesis⁵. In this model, small structural changes that are coupled to ATP hydrolysis in the catalytic domain of the myosin head are magnified by pivoting of the ~10-nm-long myosin light-chain domain, which acts as a lever arm⁵⁻⁷. The translation caused by this pivoting of the lever arm would be ~6 nm (refs 8, 9).

New techniques for manipulating single actin filaments using microneedles¹⁰ or optical traps¹¹ and optical sensors that can resolve objects at nanometre scales^{11,12} have allowed the displacement of single molecules of myosin or its subfragments to be measured directly *in vitro*¹¹⁻¹⁸. The size of displacements reported, however, has varied considerably. Some investigators have found myosin displacements of 4–10 nm (refs 11, 14–17), which are consistent with the lever-arm model. On the other hand, we have reported values of 10–25 nm (refs 18–20), which is significantly larger than the values expected from the lever-arm model. Large displacement values indicate that myosin heads may interact several times with an actin filament for each ATP used^{12,21-24}. It is essential to measure the displacement unambiguously to determine how conformational changes in the myosin head lead to force generation, and how mechanical cycles are coupled to ATP hydrolysis.

We have developed a new method for directly manipulating single myosin subfragment-1 (S1) molecules. S1 is the head region of myosin and contains the ATP- and actin-binding sites. The technique uses a scanning probe to measure the mechanical events that occur during generation of actomyosin displacements at high temporal and spatial resolution.

Nano-manipulation of single S1 molecules

The experimental arrangement is shown in Fig. 1a. S1 was biotinylated and fluorescently labelled on the regulatory light chain (RLC) at a molar ratio of >0.95:1 (fluorescent label:RLC). The RLC is a subunit of relative molecular mass 20,000 (20K) that is located near

the carboxy terminus of S1, away from the ATP- and actin-binding sites which are located on the S1 heavy chain. The scanning probe, a 5–7- μ m long ZnO crystal whisker^{25,26}, with a very sharp tip (~15 nm radius of curvature), was attached to a fine glass needle mounted on a three-dimensional piezoelectric scanner. The probe was coated with streptavidin so that single myosin heads could be captured specifically at the biotinylation site on the RLC (see Methods). The stiffness of the needles was low (0.01–0.03 pN nm⁻¹), and thus the mechanical load exerted on an S1 molecule was small (<1 pN).

Single S1 molecules captured on the tip of the scanning probe were visualized by objective-type total internal reflection fluorescence microscopy (TIRFM)²⁷, which produced clear images of single fluorophores at a high fluorescence-to-background ratio (Fig. 1b; an individual S1 molecule captured by the probe is identified by an arrowhead). The fluorescence was characterized by a single, approximately gaussian, intensity distribution and single-step photobleaching^{27,28} (data not shown), strongly indicating that the fluorescent spots were indeed due to single fluorophores. The S1 molecule was brought into contact with an actin bundle in the presence of ATP by manipulating the probe (Fig. 1a, right). We detected the displacements resulting from S1–actin interactions by measuring deflections of the needle, with subnanometre accuracy, using a differential pair of photodiodes^{12,18}.

Individual mechanical events

When S1 was not associated with an actin filament, large thermal fluctuations of the probe with an r.m.s. amplitude of ~13 nm were apparent. Upon binding of S1 to actin, the fluctuations decreased suddenly to an r.m.s. amplitude of <4.5 nm (Fig. 2a, upper trace). Motions of the probe caused by S1–actin interactions could be clearly distinguished from thermal noise by monitoring the increase in stiffness¹⁵, calculated as the reciprocal of the variance of the fluctuating probe position (Fig. 2a, lower trace). The concentration of ATP was low (0.1 or 1 μ M), leading to prolonged interactions between actin and myosin and thus enabling us to identify individual mechanical events. During S1–actin attachments, the stiffness of the probe–S1–actin linkage increased by more than tenfold to 0.2–1.5 pN nm⁻¹ at 20 °C, and by more than fivefold to

0.1–0.7 pN nm⁻¹ at 27 °C. The stiffness at 20 °C is similar to estimates of actomyosin crossbridge stiffness (0.5–2 pN nm⁻¹) in intact muscle²⁹. The high stiffness during active generation of displacements greatly improved the instrumental time resolution and signal-to-noise ratio, and allowed us to resolve elementary processes during generation of displacements (see Methods).

Displacements caused by single S1 molecules spread over a wide range of positive and negative values. The large width of this distribution is probably caused by the probe's thermal fluctuations¹⁵; if the S1 molecule attaches to an actin filament when the probe has diffused away from its equilibrium (zero) position, the start position will be displaced. The stiffness started to increase when the needle was displaced to both positive and negative positions, as shown in Fig. 2b, c, respectively. As the start positions are expected to distribute randomly around the zero position, the mean displacement over many events gives the net displacement caused by an S1 molecule¹⁵ (Fig. 2d). The mean displacements, averaged over all of the observed events, were ~13 nm with 0.1 μM ATP (*n* = 55) and ~14 nm with 1 μM ATP (*n* = 135) at 20 °C, and ~12 nm (*n* = 235) with 1 μM ATP at 27 °C. In some cases (~20%), the direction of the displacement was reversed, probably because the S1 interacted with an actin filament with the opposite orientation in the actin bundle. Therefore, the mean displacements obtained in this way should underestimate the true mean by ~20%. The durations of displacements at 0.1 and 1 μM ATP were distributed exponentially, with means of 2,400 ms (*n* = 55) and 220 ms (*n* = 135), respectively, at 20 °C. The second-order rate constant for dissociation of S1–actin by ATP, based on these mean durations, was 4–5 × 10⁶ M⁻¹ s⁻¹, which is consistent with values obtained for an actin–myosin-head complex in solution³⁰ and by optical trapping nanometry^{11,15–20}.

Stepping motion

We scrutinized the rising phase of the displacements to determine the detailed mechanism of sliding. Figure 3 shows the time course of the rising phases on an expanded timescale, in the presence of 1 μM ATP (Fig. 3a) and 0.1 μM ATP (Fig. 3b) at 20 °C, and in the presence of 1 μM ATP at 27 °C (Fig. 3c). The displacements did not take place abruptly but instead developed in a stepwise fashion. The steps at 20 °C were clearer than those at 27 °C, because the dwell time between steps was longer.

To analyse the step size for unitary and multiple displacements, we used a statistical method similar to that previously applied for the microtubule-bound molecular motor kinesin³¹ (see Methods). The start point for a step had to have a dwell time of >4 ms, and the variance of the starting position had to be no greater than double that at the final plateau position (Fig. 2b, c). This method excludes steps that have pre-attachment dwell times of <4 ms.

A histogram of pairwise distances during the rising phase of displacements was produced for each record. The data from 66 events obtained in 8 different experiments (50 events from 6 S1 molecules at 1 μM ATP and 16 events from 2 S1 molecules at 0.1 μM ATP) at 20 °C were pooled together (Fig. 4a). A new probe was used for each experiment. About 20 events were observed in each experiment. Of the 190 events, 35% (66 events) had a stiffness upon binding of >0.5 pN nm⁻¹; 65% of the events had too small a stiffness to allow us to determine the probe position reliably, and were thus not analysed—when the stiffness upon binding was small, the start positions could not be determined as reliably because of large fluctuations of the probe position. Low stiffness may have occurred when the actin bundle was not rigidly attached to the glass surface or when some actin filaments were not rigidly incorporated into the actin bundle. Data at 27 °C were also excluded because the dwell time was short and therefore the start positions could not be reliably determined. The histogram shows a peak spacing at 5.3 ± 0.2 nm (mean ± s.e.m.) and a small shoulder at twice the displacement, ~11 nm. The shoulder is probably caused by the occurrence of two sequential steps within the interval (which was 2 ms) used for calculating the distance histogram (see Methods). A small shoulder was identified near -5 nm. This shoulder may have resulted from backward steps (Fig. 3b). The peak at 5.3 nm in the histogram is statistically significant according to χ² test of goodness of fit and a non-parametric test based on a kernel-type probability density function (see Methods).

Figure 4b shows histograms of dwell times between steps. The data are well fitted by single exponentials with time constants of 3.2 and 4.8 ms in the presence of 0.1 and 1 μM ATP, respectively, indicating that the steps may have occurred stochastically. The mean velocities in the rising phase, obtained from step size divided by dwell time, are 1.7 and 1.1 μm s⁻¹ at 0.1 and 1 μM ATP, respectively, at 20 °C. These velocities are close to the maximum sliding velocities between actin and myosin at saturating concentra-

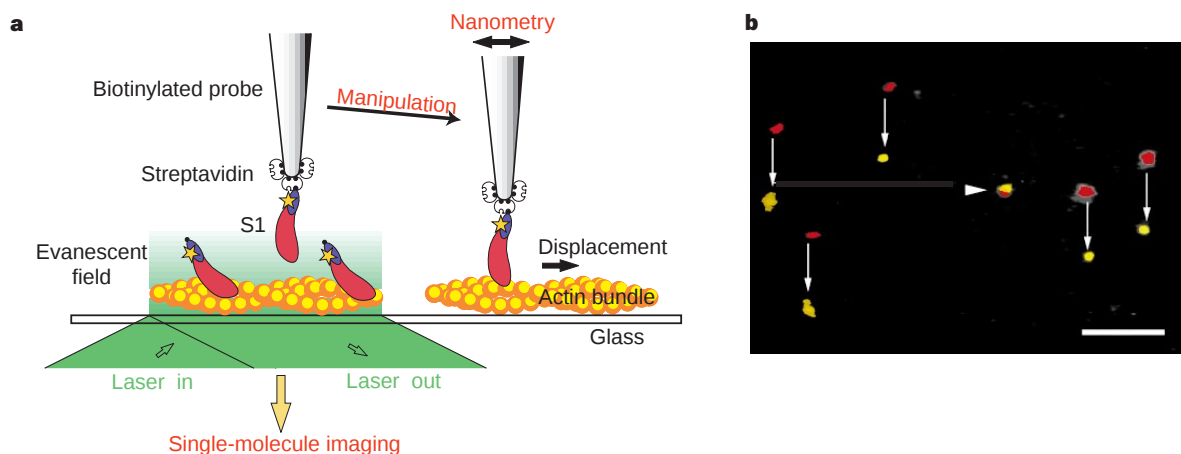


Figure 1 Direct capture and manipulation of a single S1 molecule by a scanning probe. **a**, Schematic drawing of the experiment. Left, a single S1 molecule, which was biotinylated (small black circle) and fluorescently labelled Cy3 (yellow star; molar ratio >0.95:1 Cy3:S1) at its regulatory light chain (blue), was specifically attached at its tail end through the biotin–streptavidin system to a scanning probe (ZnO whisker and glass microneedle) and observed by objective-type TIRFM²⁷. Right, the displacement produced when the S1 molecule was brought into contact with an actin bundle bound to a glass surface was determined by

measuring the position of the needle with nanometre accuracy. See text and Methods for details. **b**, Fluorescence images of single S1 molecules. The micrograph shows superimposed images of single S1 molecules either captured by the probe (arrowheads) or bound to coverslip. The red and yellow spots represent, respectively, images before and after movement of the stage by a piezoelectric actuator. The captured S1 molecule (arrowheads) did not move with the stage, but could be moved independently by piezoelectric scanners holding the needle. Calibration bar, 5 μm.

tions of ATP (>1 mM) and zero load at 20°C , and are >100 -fold larger than the velocities occurring during steady sliding at 0.1 and $1\text{ }\mu\text{M}$ ATP³². The mean velocity ($3.1\text{ }\mu\text{m s}^{-1}$) at 27°C , obtained from the slope of the rising phase of averaged events ($n = 35$), was roughly twofold larger than that at 20°C ($Q_{10} = 2.7$).

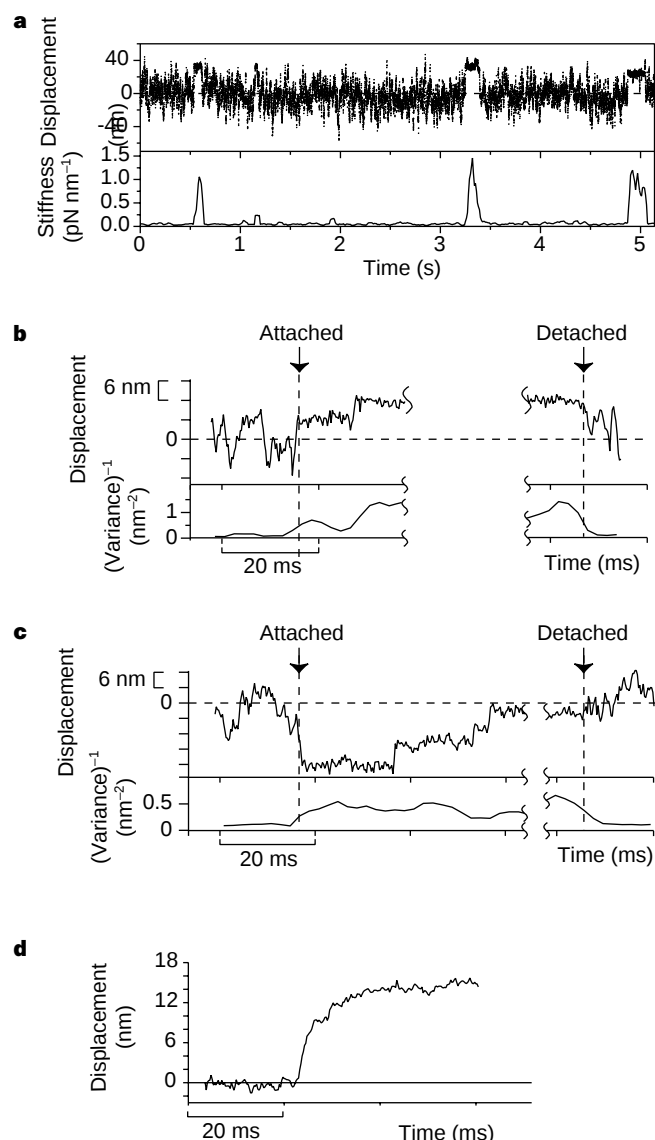


Figure 2 Displacements caused by single S1 molecules captured on the tip of the scanning probe. **a**, Upper trace, typical recordings of the displacements made by an S1 molecule. Bandwidth for recording, 2 kHz. The spring constant of the needle was 0.03 pN nm^{-1} . Lower trace, stiffness calculated at intervals of 20 ms from the variance of the probe position. **b**, **c**, Displacements due to S1-actin interactions starting at positive (**b**) and negative (**c**) positions. Upper traces, probe-position records plotted on an expanded timescale. Dotted horizontal lines show the equilibrium (zero) position of the probe. Bandwidth, 2 kHz. Lower plots indicate (variance)⁻¹. A 2-ms window was positioned every 1 ms and the variance during the window was calculated from samples at 0.25 ms intervals. In this calculation, the whole range of frequencies of probe fluctuations is not covered because the width of the window (2 ms) was small. 'Attached' and 'detached' indicate attachment and detachment of S1 to actin. **d**, Mean displacement averaged over all observed events ($n = 135$; see text). In averaging the events, we synchronized the start positions of rising phases. As the durations of rising phases varied from event to event, the end positions were not synchronized. Experiments were done in assay buffer containing an oxygen-scavenging system^{10,32} and $1\text{ }\mu\text{M}$ ATP at 20°C .

The number of steps in each event ranged from one to five, and was 2.5 steps on average (Fig. 4c); that is, the overall displacements ranged from 5.3 to 30 nm and averaged 13 nm. For individual S1 molecules, the distribution of step number ranged from one to three and one to five. The mean displacement (~ 13 nm) averaged over the events ($n = 16 + 50$, obtained in the presence of 0.1 and $1\text{ }\mu\text{M}$ ATP, respectively) selected for high stiffness at 0.1 and $1\text{ }\mu\text{M}$ ATP at 20°C was similar to that (~ 14 nm) averaged over all of the observed events ($n = 55 + 135$). This result indicates that our selection of events for analysis of the steps did not bias the distribution towards large overall displacements.

In the falling phase of the displacements (detachment of the S1 molecule from actin following ATP binding), the probe returned to the zero position with a $1/e$ time of 2–5 ms (Fig. 3a, c, last traces), which was similar to the settling time of the free needle (see Methods). No regular steps were observed either by eye or on a histogram of pairwise distances calculated during the falling phase.

Discussion

In this study, the mean displacement caused by single S1 molecules was 13 nm and displacements of up to 30 nm were observed. These values are greater than those (4–10 nm) reported previously from use of optical trapping nanometry^{11,14–17}. However, in the previous studies, because of the large variance in start positions¹⁵

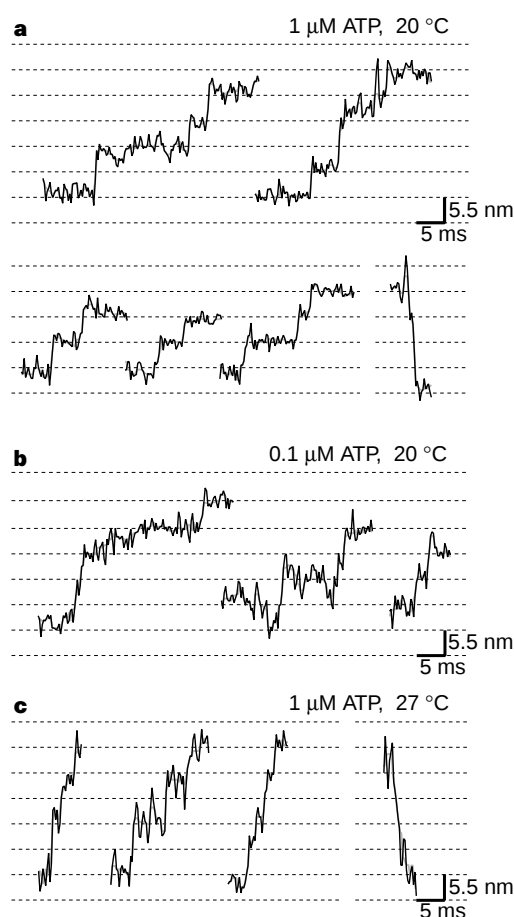


Figure 3 Steps in the rising phase of displacement records on an expanded timescale. Displacements in the presence of **a**, $1\text{ }\mu\text{M}$ ATP, 20°C ; **b**, $0.1\text{ }\mu\text{M}$ ATP, 20°C ; and **c**, $1\text{ }\mu\text{M}$ ATP, 27°C . Black lines show records passed through a 2-kHz-bandwidth low-pass filter; grey lines show the same records passed through a nonlinear median filter³¹ of rank 2. Dashed horizontal gridlines indicate spacings of 5.5 nm. The start positions of the rising phase of displacements were determined as described in the text. The last traces in **a**, **c** are records of the falling phase of displacements.

(± 30 nm; r.m.s. ~ 10 nm), only a mean displacement could be determined, rather than the distribution of individual step sizes. Thus, in previous work, some displacements could have been >10 nm. Here we have determined the size of individual displacements directly (Fig. 3). Furthermore, in our study the S1 molecule was brought close to the actin bundle, which was rigidly bound to a glass surface; thus the S1 molecule was more easily able to interact continuously with the actin filaments. In addition, the S1 molecules were specifically attached at their tail ends to the probe by an avidin–biotin system; this avoided damage to the S1 molecules caused by interaction with the probe surface. We have shown previously that the velocity of movement of actin filaments obtained with S1 molecules that are bound to a glass surface through streptavidin–biotinylated BSA, as here, is as fast as the velocity with intact myosin and several fold larger than that obtained when S1 is bound directly to the glass surface³³.

Thermal diffusion of the probe spreads the starting position of displacements, but it cannot account for the number of steps or the large displacements that occur following attachment to actin. In the presence of ADP, the mean displacement was zero and no steps were observed (data not shown). Furthermore, if the displacements resulted from thermal diffusion, the velocity in the rising phase would be proportional to the square root of the absolute temperature. The velocity, however, increased roughly twofold when the temperature was raised from 20 °C to 27 °C (Fig. 3), which is much more than would be expected from thermal diffusion.

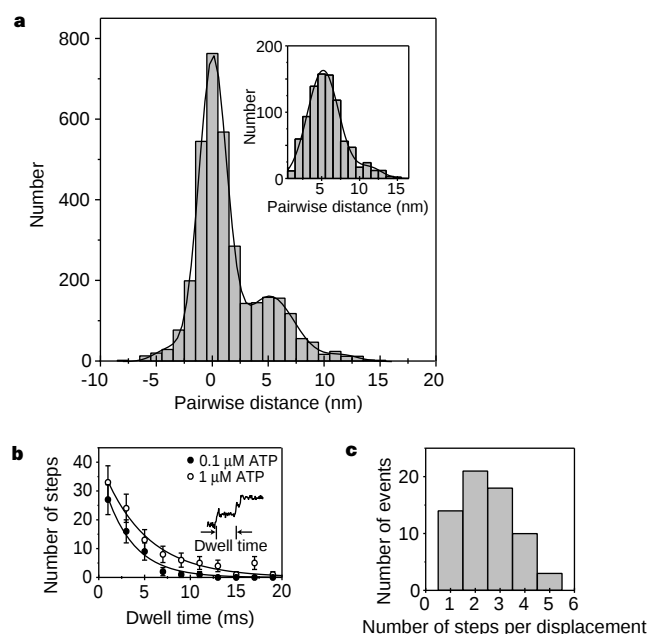


Figure 4 Statistical analysis of stepping motion. **a**, Histogram of pairwise distances in the rising phase of the displacements. The solid line is the sum of four gaussian peaks fitted to the data by least squares⁴⁹. The position and width of each gaussian peak were allowed to vary during the fitting iteration. The fit parameters for positions and widths of four gaussians are -3.8 and 1.3 nm, 0.02 and 1.3 nm, 5.3 and 2.0 nm, 11.3 and 1.6 nm, respectively. Inset, the Gaussian distribution at ~ 0 nm was subtracted from the data and the fitted curve. **b**, Histogram of the dwell times between steps, determined as described³⁵. Values are expressed as $n \pm \sqrt{n}$. Closed and open circles indicate results obtained in the presence of 0.1 and $1 \mu\text{M}$ ATP, respectively. The lines show exponential curves fitted to the data, giving time constants of 3.2 ms and 4.8 ms at 0.1 and $1 \mu\text{M}$ ATP, respectively. **c**, Histogram of the number of steps per displacement. The steps were counted by eye. Steps of $\sim 5.3 \text{ nm} \times N$ were counted as N steps. The histogram shows combined results obtained in the presence of 0.1 or $1 \mu\text{M}$ ATP at 20°C .

We did not observe displacements of more than 30 nm, even though the energy required to produce a 30 -nm displacement under the present load ($1/2 \times (30 \text{ nm})^2 \times (0.03 \text{ pN nm}^{-1}) = 14 \text{ pN} \times \text{nm}$) is $<15\%$ of the energy liberated by ATP hydrolysis ($\sim 100 \text{ pN} \times \text{nm}$)³⁴. The displacement may be limited by the helical pitch of actin. If an S1 molecule moves 30 nm along a single strand of an actin filament, the S1 would rotate by $\sim 145^\circ$ around the filament axis. Further rotation may be difficult, because of mechanical constraints at the probe or in the filament bundle.

Could the large displacements composed of successive 5.3 -nm steps be produced by multiple S1 molecules? We observed bound fluorophores rather than S1 molecules, so it is possible that the probe may have bound extra, non-fluorescent S1 molecules that were either unlabelled or labelled with photobleached fluorophores. However, this is unlikely, for several reasons. First, S1 molecules, labelled with fluorophores at a molar ratio of $>0.95:1$, were attached to the tip of the probe under safe, red-light illumination (see Methods). Illumination with green laser light was then used to check that single S1 molecules were captured (Fig. 1b). More than 90% of S1 molecules should be fluorescent, at least until this checking procedure is carried out, so most instances of capturing multiple myosins would have been detected by greater fluorescence intensity. Second, the probability of capturing more than one S1 molecule, determined by fluorescence, was $\sim 5\%$ (see Methods) and the fraction of contaminating, non-fluorescent, S1 molecules in the preparation was $<10\%$, so the likelihood that the probe bound one fluorescent S1 molecule and one or more non-fluorescent S1 molecule is $<0.5\%$. Third, if multiple steps are caused by multiple S1 molecules each producing a single step for each ATP used, then the velocity of the rising phase should have been much smaller at $1 \mu\text{M}$ ATP than the actomyosin sliding velocity at a saturating ATP concentration, and should have been lower still in the presence of $0.1 \mu\text{M}$ ATP³². This is because, at low ATP concentrations, the myosin head forms a rigid, long-lived rigor complex with actin that limits the rate of movement.

How are the steps coupled to biochemical cycles of ATP hydrolysis? During the production of one to five steps, to generate displacements of ~ 5 to ~ 30 nm (Fig. 3), the mean dwell times were 3.2 and 4.8 ms in the presence of 0.1 and $1 \mu\text{M}$ ATP, respectively, at 20°C (Fig. 4b). If a single step corresponds to each ATP used, an ATP molecule must bind to the S1 within these dwell times. The likelihood of this is ~ 3.2 ms divided by $2,400$ ms and ~ 4.8 ms divided by 220 ms in 0.1 and $1 \mu\text{M}$ ATP, respectively, where $2,400$ ms and 220 ms are the average times required for ATP binding, estimated from the plateau durations. If the hydrolysis of a single ATP molecule does correspond to a single step, and if the S1 undergoes more than two steps sequentially, more than two ATP molecules should associate repeatedly with the S1, a much less likely occurrence. The fact that the dwell time did not depend on the ATP concentration indicates that each step did not correlate with the binding of ATP. For kinesin^{35,36} and the F_1 -ATPase³⁷, each translational or rotational step is thought to require hydrolysis of an ATP molecule; for these motor proteins, the dwell times depend strongly on the ATP concentration. Thus, our results indicate that multiple elementary steps may be produced during each biochemical cycle.

The step size coincides approximately with the distance between adjacent actin monomers in one strand of an actin filament (5.5 nm; ref. 38) and myosin heads are known to bind two adjacent actin subunits during rigor⁶. A myosin head may walk along actin by using these two binding sites, without detaching from the filament. Our results do not imply, however, that the myosin head remains continuously attached to actin during movement. The probe and the needle that we used are much larger than S1, and their diffusion rate is slow. If S1 dissociates, it could move or diffuse to the adjacent actin before the probe diffuses away from the actin bundle. Thus, continuous movement along the actin filament or adjacent filaments in the bundle could involve rapid

attachment—(movement)—detachment cycles. Each step may be produced by a conformational change in the myosin head, as suggested by the lever-arm model, or by other mechanisms such as the thermal ratchet (see, for example, ref. 39). It is also possible that conformational changes within the actin filament are important in force generation^{40,41}.

To allow for the production of multiple steps for each molecule of ATP split, chemical energy from ATP hydrolysis might be stored in the myosin head or in the actin filament and released gradually during successive actomyosin interactions. This idea challenges the widely accepted view that force generation is directly coupled to release of bound ligands⁴². We have shown recently that the myosin head can attach to actin and generate force for a considerable time (>100 ms) after the release of bound nucleotide¹⁹. In some other proteins, ligand-induced conformations, termed imprints or protein memory, are retained after removal of the ligand⁴³. Our results indicate that the myosin head may be able to store energy from ATP hydrolysis and later release it productively in several packets of work. □

Methods

Proteins. S1 was obtained by papain digestion of chicken skeletal muscle myosin⁴⁴ and purified by HPLC gel filtration (Superose-6, Pharmacia). Contamination with heavy meromyosin and myosin was below the limit for detection (<0.5%). Biotinylated and fluorescently labelled S1 was obtained by exchanging the RLC for a biotinylated and labelled light chain as described²⁷. A recombinant fusion protein, consisting of biotin-dependent transcarboxylase (BDTC), a peptide biotinylated in *Escherichia coli*, and chicken gizzard myosin RLC, was expressed in *E. coli* strain JM109 and purified³³. The single cysteine residue of RLC was labelled with an orange fluorescing cyanine reagent, Cy3-maleimide, which was prepared by coupling monofunctional Cy3 N-hydroxysuccinimide ester (Amersham) with N-[2-(1-piperazinyl)ethyl]-maleimide (Dojindo Laboratories, Japan), and Cy3-BDTC-RLC was exchanged for endogenous RLC²⁷. The Cy3:S1 molar ratio was estimated spectrophotometrically from the suspension in solution to be >0.95 using the extinction coefficients $0.83 \text{ (mg/ml)}^{-1} \text{ cm}^{-1}$ at 280 nm for S1 (ref. 44), and $150,000 \text{ M}^{-1} \text{ cm}^{-1}$ at 552 nm for Cy3 (product data sheet). By measuring the absorption of Cy3 in the absence and presence of a twofold molar excess of BDTC-RLC, we found that the extinction coefficient for Cy3 was unchanged (it changed by <2%) upon binding to RLC. The contribution of the absorbance of BDTC to that of S1 was negligible (<1%). To determine whether the suspension was a mixture of S1 molecules that bound none, one and more than one fluorescent Cy3 molecule, we measured the fluorescence intensity and photobleaching of individual S1 molecules bound to a glass surface by TIRFM²⁷. The fluorescence intensity showed a single, approximately gaussian-shaped distribution. 169 out of 179 fluorescent spots, corresponding to S1 molecules, were photobleached in one step, and the other 10 spots were photobleached in two steps. Thus 6% of the S1 molecules apparently bound two dye molecules; this percentage is an upper limit, because two S1 molecules may occasionally have been in close proximity on the coverslip.

Actin⁴⁵ and α -actinin⁴⁶ were obtained from rabbit skeletal muscle and chicken gizzard, respectively. Actin bundles of ~100 nm in width were prepared by dialysis of G-actin and α -actinin in a molar ratio of 10:1 against the assay buffer (25 mM KCl, 5 mM MgCl_2 , 20 mM HEPES, pH 7.8), and then fluorescently labelled with BODIPY FL phalloidin (Molecular Probes).

Probe. ZnO whisker crystals (a gift from Central Research Laboratories, Matsushita Electric Industrial Co, Ltd) were amino-silanized as described⁴⁷, biotinylated in N,N-dimethylformamide containing 1 mM 5-[5-(N-succinimidylloxycarbonyl)pentylamido]hexyl-D-biotinamide (biotin-(AC₅)₂-Osu, Dojindo Laboratories) overnight at room temperature, washed several times with ethanol and stored in ethanol. A biotinylated ZnO whisker was glued with epoxy resin under a binocular microscope to the end of a very fine glass microneedle, ~100 μm long and ~0.3 μm in diameter¹⁸. The bending stiffness (spring constant) of the needles was calibrated by measuring the mean square of their thermal fluctuations using the principle of energy equipartition¹⁸, or by cross-calibration against stiffer calibrated needles¹⁸. The compliance added to the actomyosin-probe-needle system by torsion of the glass needle was

calculated to be $0.5\text{--}1 \text{ nm pN}^{-1}$. When the probe was attached directly to a glass surface, fluctuations of the probe were reduced below the detector noise (~0.3 nm r.m.s.), indicating that the ZnO whisker and its epoxy attachment were rigid.

Observation and manipulation of single S1 molecules. Cy3-labelled BDTC-S1 (Cy3-BDTC-S1) was visualized under an improved fluorescence microscope (IMT-2, Olympus) equipped for objective-type TIRFM²⁷ and including a nanometre-piconewton measurement system^{12,18}. Fluorescence intensity and photobleaching were measured under the same conditions as described²⁷. Both the Cy3-labelled S1 and BODIPY-FL-labelled actin bundles were excited by a frequency-doubled Nd:YAG laser ($\lambda = 532 \text{ nm}$)⁴⁸. A solution containing the actin bundles was applied to a glass surface that had been cleaned and then coated with α -actinin (0.1 mg ml^{-1}). Unbound actin bundles were washed out with the assay buffer. $\sim 10^{-16} \text{ mol}$ of Cy3-BDTC-S1 in assay buffer without ATP was then applied to the coverslip. After 1 min of incubation, unbound Cy3-BDTC-S1 was washed out. The surface was then washed with assay buffer containing 0.5% 2-mercaptoethanol and an oxygen-scavenging system^{10,32} to reduce photobleaching. S1 molecules bound to actin bundles fixed on the glass surface were observed briefly (for ~1 s) under 532-nm illumination, a short period compared with the average photobleaching time on actin bundles (~60 s).

The streptavidin-coated probe was then positioned above the surface of a coverslip and scanned over a $1 \mu\text{m} \times 1 \mu\text{m}$ area to capture a single S1 molecule through the streptavidin-biotin bond. Contact of the probe with the surface was monitored by observing thermal fluctuations of the needle, under He-Ne laser ($\lambda = 632.8 \text{ nm}$) illumination. The green laser was turned off during this part of the procedure to suppress photobleaching. Then the fluorescence at the tip of the probe was observed under 532-nm illumination, as described above, to determine, according to the fluorescence intensity and single-step photobleaching²⁷ whether a single S1 molecule had bound. During the time required (~5 min) to complete all experimental manipulations before checking for capture of a single S1, under the red-light and room-lighting conditions used, the number of photobleached Cy3 molecules bound to S1 attached to actin bundles on a glass surface was ~2% (out of 1,000 observed fluorescent spots). If no S1 molecule was captured, the probe was replaced and this procedure was repeated until an S1 molecule was successfully captured. It was rare that more than one S1 molecules was captured, as determined by fluorescence (1 of 20 trials in which S1 molecule(s) were captured). At the S1 concentration, solution and illumination conditions, and duration of the experiments used, we did not observe binding of more than two molecules to the tip of the probe. Damage of an S1 molecule by physical contact with the probe during this process is likely to be small because the needle was very flexible. The stiffness of the needle was >1,000-fold less than that of the cantilevers generally used for scanning probe microscopy (>0.1 nN nm^{-1}). The probe and captured S1 were manipulated in three dimensions, with subnanometre accuracy, by piezoelectric scanners.

Displacement measurements. After the S1 molecule was captured on the probe tip, ATP was added to the medium. The S1 was then brought into contact with a fresh actin bundle bound to a different region of the same coverslip (Fig. 1a). The angle between the needle and the actin bundle was set at ~90°. The interaction between the single S1 molecule and the actin filaments was monitored by measuring deflections of the glass microneedle with subnanometre resolution using a split-photodiode and the 532-nm laser¹⁸. The response time (rise time from 0 to 63% of a free needle was 2–5 ms, as determined from the corner frequency ($f_c = 30\text{--}60 \text{ Hz}$) of the power spectrum of the needle's thermal fluctuations: the response time (τ) = $1/(2\pi f_c)$). During generation of displacements, however, the stiffness of the probe-S1-actin linkage (k_{s1}) increased, so that τ should be greatly improved according to the equation $\tau = \zeta/(k_{s1} + k_{\text{needle}})$, where ζ is the friction coefficient. τ is calculated to be < 0.4 ms at $k_{s1} > 0.5 \text{ pN nm}^{-1}$. This response time was sufficient to observe steps occurring on a millisecond timescale. Thermal fluctuations of a free needle were large (r.m.s. amplitude ~13 nm), but they decreased during generation of displacements because of the increased stiffness (Fig. 2). This effect can be quantified using the principle of energy equipartition, $\langle x^2 \rangle^{1/2} = (k_B T/(k_{s1} + k_{\text{needle}}))^{1/2}$, where k_B is Boltzmann's constant and T is the absolute temperature. When k_{s1} is $> 0.5 \text{ pN nm}^{-1}$, $\langle x^2 \rangle^{1/2}$ is <2.5 nm, which is small enough not to obscure the stepping motions.

Data acquisition and analysis. Probe position data were acquired at 24 kHz by a DAT recorder (RD-125T, TEAC, Japan), and analysed offline by using DADiSP (DSP Development Corp.), LabVIEW (National Instruments) and Origin (Microcal, Inc.) software. The acquired data were passed through a digital recursive Chebyshev type I filter of bandwidth 2 kHz. Pairwise distances were obtained by a method used previously for determining the step size of kinesin³¹, with some modifications. Before the pairwise-distance analysis, the rising phases of the displacements were further filtered by a nonlinear median filter of rank 2 (ref. 31). The traces during the rising phase were shifted by 2 ms and the distances between the original and shifted traces were calculated every 0.25 ms. Reliability of the signal-processing and statistical analysis was checked by the following observations. When a single S1 molecule attached to the probe was bound to an actin bundle on a glass surface in the absence of ATP and the stage was moved smoothly at a velocity similar to that occurring in the rising phase of the displacements, the pairwise-distance histogram showed a single gaussian peak and no auxiliary peaks at a spatial periodicity near 5.3 nm. As the stage could not be moved in 5.3-nm steps rapidly enough to simulate the active steps, simulated data with stochastic 5.3-nm steps and gaussian noise were processed in the same way as experimental data. The distance histogram was indistinguishable from that shown in Fig. 4a.

Statistics. The sum of several gaussian distributions was fitted to the data by the least-squares method using the Levenberg–Marquardt algorithm⁴⁹. The fitted curve was tested by the χ^2 test of goodness of fit (significance level, $\alpha = 0.005$). The data were further tested by a non-parametric test based on the kernel-type probability-density function (PDF)⁵⁰. The estimator of the PDF is defined as:

$$\hat{f}(x : h) = (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

where n is the number of the data, X_i is the i th datum and h is the bandwidth that satisfies the relationship $h \propto n^{1/5}$. The standard normal distribution function has been used as the kernel $K(\cdot)$. The sign of the differential coefficient of the PDF shows the increase or the decrease of the PDF which indicates the position of the peaks within the PDF.

Received 17 March; accepted 1 December 1998.

- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin: DNase I complex. *Nature* **347**, 37–44 (1990).
- Rayment, I. et al. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50–58 (1993).
- Huxley, H. E. The mechanism of muscular contraction. *Science* **164**, 1355–1366 (1969).
- Huxley, A. F. & Simmons, R. M. Proposed mechanism of force generation in striated muscle. *Nature* **233**, 533–538 (1971).
- Spudich, J. A. How molecular motors work. *Nature* **372**, 515–518 (1994).
- Rayment, I. et al. Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
- Fisher, A. J. et al. X-ray structures of myosin motor domain of *Dictyostelium discoideum* complexed with MgADP·BeF₃ and MgADP·AlF₄. *Biochemistry* **34**, 8960–8972 (1995).
- Cooke, R. Actomyosin interaction in striated muscle. *Physiol. Rev.* **77**, 671–697 (1997).
- Howard, J. Molecular motors: structural adaptations to cellular functions. *Nature* **389**, 561–567 (1997).
- Kishino, A. & Yanagida, T. Force measurements by micromanipulation of a single actin filament by glass needles. *Nature* **334**, 74–76 (1988).
- Finer, J. T., Simmons, R. M. & Spudich, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
- Ishijima, A., Doi, T., Sakurada, K. & Yanagida, T. Sub-piconewton force fluctuations of actomyosin *in vitro*. *Nature* **352**, 301–306 (1991).
- Ishijima, A. et al. Single-molecule analysis of the actomyosin motor using nano-manipulation. *Biochem. Biophys. Res. Commun.* **199**, 1057–1063 (1994).
- Miyata, H. et al. Stepwise motion of an actin filament over a small number of heavy meromyosin molecules is revealed in an *in vitro* motility assay. *J. Biochem. (Tokyo)* **115**, 644–647 (1994).
- Molloy, J. E., Burns, J. E., Kendrick-Jones, J., Tregear, R. T. & White, D. C. S. Movement and force produced by a single myosin head. *Nature* **378**, 209–212 (1995).
- Guilford, W. H. et al. Smooth muscle and skeletal muscle myosins produce similar unitary forces and displacements in the laser trap. *Biophys. J.* **72**, 1006–1021 (1997).

- Mehta, A. D., Finer, J. T. & Spudich, J. A. Detection of single-molecule interaction using correlated thermal diffusion. *Proc. Natl Acad. Sci. USA* **94**, 7927–7931 (1997).
- Ishijima, A. et al. Multiple- and single-molecule analysis of the actomyosin motor by nanometer-piconewton manipulation with a microneedle: unitary steps and forces. *Biophys. J.* **70**, 383–400 (1996).
- Ishijima, A. et al. Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* **92**, 161–171 (1998).
- Tanaka, H., Ishijima, A., Honda, M., Saito, K. & Yanagida, T. Orientation dependence of displacements by a single one-headed myosin relative to the actin filament. *Biophys. J.* **75**, 1886–1894 (1998).
- Yanagida, T., Arata, T. & Oosawa, F. Sliding distance of actin filament induced by a myosin crossbridge during one ATP hydrolysis cycle. *Nature* **316**, 366–369 (1985).
- Harada, Y., Sakurada, K., Aoki, T., Thomas, D. D. & Yanagida, T. Mechanochemical coupling in actomyosin energy transduction studied by *in vitro* movement assay. *J. Mol. Biol.* **216**, 49–68 (1990).
- Higuchi, H. & Goldman, Y. E. Sliding distance between actin and myosin filaments per ATP molecule hydrolysed in skinned muscle fibres. *Nature* **352**, 352–354 (1991).
- Lombardi, V., Piazzesi, G. & Linari, M. Rapid regeneration of the actin-myosin power stroke in contracting muscle. *Nature* **355**, 638–641 (1992).
- Kitano, M., Hamabe, T., Maeda, S. & Okabe, T. Growth of large tetrapod-like ZnO crystals. *J. Crystal Growth* **102**, 965–973 (1990).
- Kado, H., Yokoyama, K. & Tohda, T. Atomic force microscopy using ZnO whisker tip. *Rev. Sci. Instrum.* **63**, 3330–3332 (1992).
- Tokunaga, M., Kitamura, K., Saito, K., Iwane, A. H. & Yanagida, T. Single molecule imaging of fluorophores and enzymatic reactions achieved by objective-type total internal reflection fluorescence microscopy. *Biochem. Biophys. Res. Commun.* **235**, 47–53 (1997).
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
- Huxley, A. F. & Tideswell, S. Filament compliance and tension transients in muscle. *J. Muscle Res. Cell Motil.* **17**, 507–511 (1996).
- Wolejda, R. C., Curtin, N. A. & Homsher, R. *Energetic Aspects of Muscle Contraction* Ch. 3 (Academic, London, 1985).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Harada, Y., Noguchi, A., Kishino, A. & Yanagida, T. Sliding movement of single actin filaments on one-headed myosin filaments. *Nature* **326**, 805–808 (1987).
- Iwane, A. H., Kitamura, K., Yokunaga, M. & Yanagida, T. Myosin subfragment-1 is fully equipped with factors essential for motor function. *Biochem. Biophys. Res. Commun.* **230**, 76–80 (1997).
- Bagshaw, C. R. *Muscle Contraction* (Chapman & Hall, London, 1993).
- Kojima, H., Muto, E., Higuchi, H. & Yanagida, T. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* **73**, 2012–2022 (1997).
- Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
- Yasuda, R., Noji, H., Kinosita, K. Jr & Yoshida, M. F₁-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps. *Cell* **93**, 1117–1124 (1998).
- Huxley, H. E. & Brown, W. The low-angle x-ray diagram of vertebrate striated muscle and its behavior during contraction and rigor. *J. Mol. Biol.* **30**, 383–434 (1967).
- Astumian, R. D. Thermodynamics and kinetics of a brownian motor. *Science* **276**, 917–922 (1997).
- Egelman, E. H., Francis, N. & DeRosier, D. J. F-actin is a helix with a random variable twist. *Nature* **298**, 131–135 (1982).
- Yanagida, T., Nakase, M., Nishijima, K. & Oosawa, F. Direct observation of motion of single F-actin filaments in the presence of myosin. *Nature* **307**, 58–60 (1984).
- Lynn, R. W. & Taylor, E. W. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* **10**, 4617–4624 (1971).
- Klibanov, A. M. Enzyme memory. What is remembered and why? *Nature* **374**, 596 (1995).
- Margossian, S. S. & Lowey, S. Preparation of myosin and its subfragments from rabbit skeletal muscle. *Methods Enzymol.* **85**, 55–71 (1982).
- Spudich, J. A. & Watt, S. The regulation of rabbit skeletal muscle contraction. I. Biochemical studies of the interaction of the tropomyosin-troponin complex with actin and the proteolytic fragments of myosin. *J. Biol. Chem.* **246**, 4866–4871 (1971).
- Craig, S. W., Lancashire, C. L. & Cooper, J. A. Preparation of smooth muscle α -actinin. *Methods Enzymol.* **85**, 316–321 (1982).
- Tokunaga, M., Aoki, T., Hiroshima, M., Kitamura, K. & Yanagida, T. Subpiconewton intermolecular force microscopy. *Biochem. Biophys. Res. Commun.* **231**, 566–569 (1997).
- Saito, K., Tokunaga, M., Iwane, A. H. & Yanagida, T. Dual color microscopy of single fluorophores bound to myosin interacting with fluorescently-labelled actin using anti-Stokes fluorescence. *J. Microsc.* **188**, 255–263 (1997).
- Bevington, P. R. & Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1992).
- Wand, M. P. & Jones, M. C. *Kernel Smoothing* (Chapman & Hall, London, 1995).

Acknowledgements. We thank T. Funatsu, K. Saito, H. Higuchi, A. Ishijima and H. Kojima for technical suggestions; Y. Ishii and other colleagues of the ERATO project and Osaka University for valuable discussions; A. F. Huxley, Y. E. Goldman, F. Brozovich, C. R. Bagshaw, J. E. Molloy, A. D. Mehta, R. D. Vale and J. West for critically reading the manuscript; S. Kimura for instructions on preparing α -actinin; M. Taniguchi for advice on statistical analysis; and M. Kitano, H. Kado and H. Ogawa for the gift of the ZnO whiskers. This work was partially supported by JSPS Research Fellowships for Young Scientists (K.K.).

Correspondence and requests for materials should be addressed to T.Y. (e-mail: yanagida@phys1.med.osaka-u.ac.jp).

The thermal imprint of galaxy formation on X-ray clusters

Trevor J. Ponman*, Damian B. Cannon* & Julio F. Navarro†

* School of Physics & Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

† Department of Physics & Astronomy, University of Victoria, Victoria BC V8P 1A1, Canada

It is widely believed that structure in the Universe evolves hierarchically—fluctuations in the primordial distribution of matter, amplified by gravity, collapse and merge to form progressively larger systems, culminating in the clusters of galaxies that are observed today. But the observed structure and evolution of X-ray-emitting clusters of galaxies seems to be at odds with this picture¹. In particular, clusters and groups with relatively few galaxies, as well as most distant clusters, are substantially fainter in X-rays than predicted by models of hierarchical formation. Here we show that these discrepancies arise because the entropy of the hot diffuse intracluster gas near the centre of the cluster is higher than can be explained by gravitational collapse alone. We argue that the excess entropy is a relic of the energetic winds generated by supernovae in the forming galaxies. These winds also enriched the intracluster medium with elements heavier than helium. We show that such a process can account for the observed effects only if the intracluster medium is heated at modest redshifts ($z \lesssim 2$) but before the final collapse into a cluster structure, indicating that the formation of galaxies precedes that of clusters and that most clusters have been assembled very recently.

Recent numerical work has unveiled a remarkable similarity in the structure of galaxy clusters formed through hierarchical clustering: properly scaled, the dark-matter density profiles of virialized clusters are nearly identical². The scaling prescription is simple. Coeval clusters have approximately the same characteristic density³, $M/r^3 = \text{constant}$, where M and r are the characteristic mass and radius of a cluster. The mass profiles of all clusters are then similar once radii are scaled to a 'virial' radius, $r_v \propto (GM/r)^{1/2}$. Baryons in galaxy clusters contribute a minority of the total mass, and are found in a hot, diffuse, X-ray-bright intracluster medium (ICM) at a temperature, T , which indicates the depth of the potential well; $T \propto GM/r \propto r_v^{-2}$. If gas traces mass, scaled X-ray surface brightness profiles are also expected to be similar. Furthermore, a simple scaling is expected between X-ray luminosity, L_X , and temperature. Assuming that the emission is dominated by bremsstrahlung, $L_X \propto M_{\text{gas}}^2 R^{-3} T^{1/2}$, or $L_X \propto f_{\text{gas}}^2 T^2$, where $f_{\text{gas}} = M_{\text{gas}}/M$ is the gas mass fraction.

X-ray clusters deviate substantially from this simple scaling, and the observed $L_X:T$ relation⁴ is considerably steeper than T^2 . Two main interpretations are usually considered: either the gas fraction is systematically higher in hotter clusters⁵, or the spatial distribution of the gas deviates systematically from the similarity profile in clusters of different temperature. A simple example of the latter is provided by the 'preheated ICM' model^{1,6}, wherein the entropy of the gas in all clusters is raised at early times to levels comparable to those reached through gravitational collapse. This entropy 'floor' imposes a maximum central density for the ICM which increases roughly as $T^{3/2}$ and therefore affects cooler clusters more severely, reducing their X-ray luminosities and steepening the $L_X:T$ relation relative to the similarity scaling.

To investigate the nature of this similarity-breaking, we have extracted X-ray surface brightness profiles from ROSAT observations of 25 systems spanning a factor of ~ 15 in temperature, from rich clusters to small groups. The sample is restricted to systems with reasonably regular X-ray morphology and good estimates of X-

ray temperature from the Ginga or ROSAT satellites. The profiles are then scaled so that they would coincide if obeying similarity, and are shown in Fig. 1. (Details of the scaling procedure are given in Fig. 1 legend.) Profiles are colour-coded according to system temperature. Systems with $T > 4$ keV (shown in purple) have on average similar profiles and show no strong trend with temperature, whilst cooler systems (green and red) have profiles that become progressively shallower and less centrally concentrated. This is especially noticeable outside the centre, where cooling flows dominate and the profiles of most systems steepen ($r \leq 0.05 r_v$). Deviations in the scaled profiles persist out to radii where cooling timescales greatly exceed the age of the Universe, so it seems unlikely that radiative effects are responsible for this result.

The trends in Fig. 1 are clearly inconsistent with simple variations in overall baryon fraction (which would result in profiles similar in shape, but shifted vertically with T), but are in good agreement with the predictions of the 'preheated ICM' model described above. However, one can imagine other effects which would lead to a reduction in central gas densities. Possibilities include non-thermal pressure support (for example, by magnetic fields) or a failure in the similarity of the underlying gravitational potentials. The unmistakable signature of preheating is a rise in the entropy of the gas. Magnetic support, for example, would in fact reduce the entropy, as it would weaken the shocks which heat the gas as the cluster potential develops. To explore this, we have fitted simple isothermal models to the X-ray emission profiles in order to derive profiles of the 'entropy', $S(r) = T/n_e^{2/3}$, where n_e is the electron density, for all systems. The fits exclude any central excess arising from a cooling flow, and employ the widely used isothermal β -model⁷ to derive $n_e(r) = n_e^0 [1 + (r/r_{\text{core}})^2]^{-3\beta/2}$.

Figure 2 shows the entropy of the gas at a fiducial radius of $0.1 r_v$, just outside the cooling-flow-dominated region. If the gas profiles were similar, the entropy at a fixed fraction of r_v would be directly proportional to the temperature (solid line). Cool ($T < 4$ keV) clusters clearly have entropies higher than achievable through gravitational collapse alone. The presence of large cores in the entropy profiles of a small sample of galaxy groups and clusters has been noted previously⁸. Here we see that this is part of a systematic trend whereby low-mass systems progressively depart from the self-similar scaling, and that at $T \sim 1$ keV the entropy appears to converge to a characteristic floor value of $\sim 100 h^{-1/3}$ keV cm².

We now consider what processes are potentially able to establish this trend. Cooling flows will remove low-entropy gas from the centre, allowing higher-entropy material to flow inwards. However, an effect of the magnitude seen can only be obtained if the amount of gas removed by the cooling flow is a substantial fraction of the ICM, which seems unlikely⁹. Moreover, substantial loss of gas to cooling flows, or large systematic variations in galaxy formation efficiency, are inconsistent with the fact that the gas fraction within r_v , and the metallicity of the ICM, are roughly independent of the temperature of the system^{10,11}. It is clear from the observed metal content of the ICM that galaxies have 'polluted' their surroundings with substantial quantities of metal-rich gas. The strong galactic winds responsible for this pollution are thus the most natural mechanism available to heat the ICM. In hot clusters, the gravitational collapse of the system generates entropies in excess of the floor value established by galaxy winds, but in cooler systems it is preserved during collapse, and prevents the gas from collapsing to high central densities^{12,13}.

The energy requirements of the preheating mechanism can be assessed by considering the Coma cluster, the best studied nearby rich cluster, although the argument applies equally to smaller systems. The total stellar mass in galaxies in Coma is $\sim 10^{13} h^{-1} M_\odot$ (where $M_\odot$ is the solar mass) and therefore the total energy available from supernovae is $\sim 10^{62} h^{-1}$ erg (assuming 1 supernova event per $100 M_\odot$ of stars formed). As the gas mass in the ICM is $\sim 5.5 \times 10^{13} h^{-5/2} M_\odot$, these supernovae could raise the ICM tem-

perature by $\sim 0.4h^{3/2}$ keV. Hence supernovae could only raise the entropy, $Tn_e^{-2/3}$, to the observed floor value of $\sim 100 h^{-1/3}$ keV cm² if $n_e < 2.5 \times 10^{-4} h^{1/4}$ cm⁻³, a value well below the electron density at our fiducial radius of $r = 0.1r_v$ in Coma¹⁴. In other words, dumping all of the available supernova energy into the ICM of Coma today would be insufficient to heat the denser inner regions to the desired entropy floor. The problem is even more acute at higher redshift, because the density of the Universe, and that of all collapsed clumps, is higher then. This simple analysis implies that the gas in the ICM cannot be in collapsed, overdense systems at the time of preheating. The formation of galaxies must therefore precede the collapse of the cluster as a whole, validating the basic tenet of the hierarchical clustering paradigm.

When did this heating occur? Assuming that the gas is uniformly distributed throughout the Universe at the time of preheating, the density limit gives a strict upper limit on the preheating redshift, $z_{\text{preh}} \lesssim 10$. In practice, density variations and reductions in heating efficiency reduce this redshift limit. An entropy of $100h^{-1/3}$ keV cm² is approximately the value baryons would have at present if they were spread uniformly throughout the Universe at the mean density inferred from models of cosmic nucleosynthesis¹⁵ and heated to a temperature of 30,000 K. The entropy of the gas responsible for the Lyman- α 'forest' observed in quasar spectra (the dominant baryonic component of the Universe at $z \approx 2$)¹⁶ is almost an order of magnitude smaller, as the temperatures are of the same order but the density of the Universe at $z \approx 2$ is a factor of ~ 30

greater. If, as expected, winds from forming galaxies affect cluster and field environments alike, observations of the Lyman- α forest imply that $z_{\text{preh}} \lesssim 2$, consistent with the low redshift at which the global star-formation rate in the Universe appears to have peaked¹⁷.

We can also estimate the typical propagation velocity of the winds, v_w , by noting that one galaxy can perturb a volume of order $(4\pi/3)(v_w/100 \text{ km s}^{-1})^3 h^{-3} \text{ Mpc}^3$ in a Hubble time. Because the mass in the Coma cluster has been collected from a volume a few hundred times this, whilst its optical luminosity is approximately that of 200 galaxies like the Milky Way, most of the ICM in Coma could have been preheated by winds blowing at $v_w \gtrsim 100 \text{ km s}^{-1}$. Such velocities lie comfortably within the range of winds observed in galaxies which are actively forming stars¹⁸.

The results presented above demonstrate the strong influence that forming galaxies have on their surroundings. In the case of rich clusters, which account for $\sim 10\%$ of galaxies, preheating by galaxy winds will substantially modify the evolution of the X-ray luminosity and temperature functions¹⁹, and will result in a luminosity-temperature relation which is independent of redshift^{2,6}, in good agreement with current compilations of X-ray data²⁰. In poor clusters and groups, the entropy floor will have more dramatic effects, and may reduce the baryon fraction within the virial radius below the universal value²¹.

For the majority of galaxies which, like the Milky Way, are not located in virialized groups or clusters, preheating will largely prevent gas from collapsing into their potential wells, leaving the

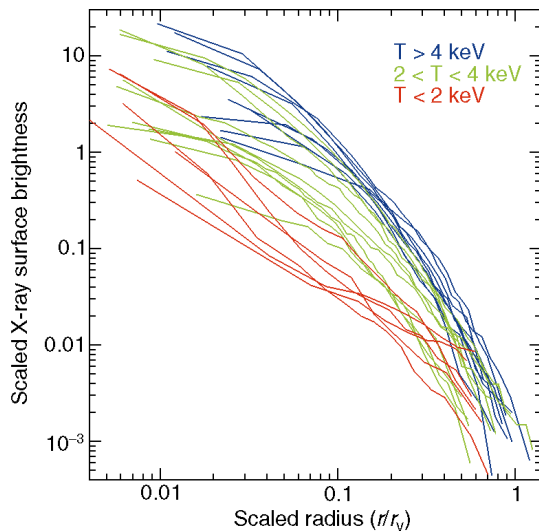


Figure 1 Scaled X-ray surface-brightness profiles overlaid to show departures from similarity in galaxy systems of different temperatures. ROSAT Position Sensitive Proportional Counter (PSPC) images in the 0.4–2 keV band of the 25 systems (A262, A478, A496, A548S, A665, A1060, A1413, A1651, A1689, A1795, A1837, A2142, A2163, A2199, A2218, AWM4, AWM7, MKW3, MKW4, HCG62, HCG94, HCG97, NGC533, NGC2300 and NGC4261) were extracted from the LEDAS public archive at Leicester University, corrected²³ for the effects of vignetting and obscuration by detector support structures, and background flux subtracted. Point sources were removed, and X-ray surface-brightness profiles constructed relative to the centroid of the remaining diffuse emission (making due correction for areas removed). Mean temperatures for each system are taken from Ginga and ROSAT observations^{24–28}. These are used to compute virial radii⁹, $r_v = 0.57(T/\text{keV})^{1/2} h^{-1} \text{ Mpc}$. The X-ray profiles have been converted from ROSAT count rates to bolometric X-ray intensities using a hot plasma code²⁹, and have been scaled by $T^{1/2}(1+z)^{3/2} \Lambda(T)$ to allow for self-similar scaling with mass and cosmic density evolution ($\Lambda(T)$ is the temperature dependence of the hot plasma emissivity), and by $(1+z)^{-4}$ to correct for cosmological dimming. The profiles are colour-coded by system temperature. Similarity appears to hold for hot ($T > 4$ keV) systems, but large systematic deviations are noticeable in cooler systems.

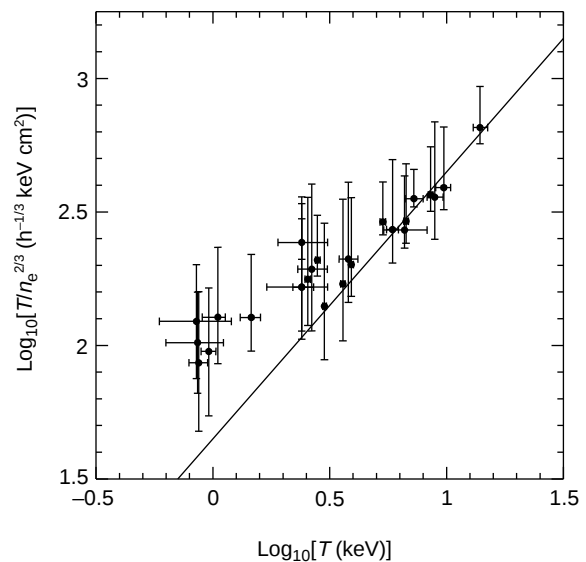


Figure 2 The gas 'entropy' at a fiducial radius $r = 0.1r_v$ as a function of temperature for the 25 systems in our sample. The entropy is defined as $S = T/n_e^{2/3}$, where T is the mean temperature and n_e is the electron number density. Entropies were computed by fitting isothermal- β models⁷ to the surface-brightness profiles shown in Fig. 1. Error bars indicate 90% confidence levels in temperature, and span the variation in entropy from $0.05 r_v$ to $0.2 r_v$. The solid line shows the similarity relation obtained from numerical simulations including non-radiative gas dynamics³⁰, $S(0.1r_v) \approx 45(T/\text{keV})(f_{\text{gas}}/0.06)^{-2/3} h^{-4/3} \text{ keV cm}^2$, where Hubble's constant is $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$. The numerical results depend only weakly on the assumed cosmological model, and provide a very good match to the central entropy of hot ($T > 4$ keV) systems for $f_{\text{gas}} \approx 0.06 h^{-3/2}$. However, poor clusters and groups have apparently been heated to higher entropy than achievable through gravitational collapse.

bulk of the baryons in the Universe in a currently undetectable high-entropy sea²². Inhibiting the collapse of baryons into galaxies after preheating robs them of their main source of fresh fuel and may explain the precipitous drop¹⁷ in the global star-formation rate of the Universe since $z \approx 1$. In this model, most baryons today are distributed between galaxies at temperatures $T \approx 30,000\delta^{2/3}$ K, where δ is the gas overdensity relative to the universal average. This explains why most galaxies lack the prominent X-ray haloes observed in rich clusters. In the haloes of galaxies like the Milky Way, where the virial temperature is ~ 0.1 keV, gas densities cannot exceed a few times the universal average, and would only be detectable through absorption studies of highly ionized species of elements such as Si in the ultraviolet spectra of background sources. Only in 'hotter' systems, such as massive elliptical galaxies, and galaxy groups and clusters, can baryons achieve densities high enough to shine brightly and be easily detected. $\square$

Received 17 July; accepted 20 October 1998.

- Kaiser, N. Evolution of clusters of galaxies. *Astrophys. J.* **383**, 104–111 (1991).
- Navarro, J. F., Frenk, C. S. & White, S. D. M. Simulations of X-ray clusters. *Mon. Not. R. Astron. Soc.* **275**, 720–740 (1995).
- Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
- White, D. A., Jones, C. & Forman, W. An investigation of cooling flows and general cluster properties from an X-ray image deprojection analysis of 207 clusters of galaxies. *Mon. Not. R. Astron. Soc.* **292**, 419–467 (1997).
- David, L., Arnaud, K. A., Forman, W. & Jones, C. Einstein observations of the Hydra-A cluster and the efficiency of galaxy formation in groups and clusters. *Astrophys. J.* **356**, 32–40 (1990).
- Evvard, A. E. & Henry, J. P. Expectations for X-ray cluster observations by the ROSAT satellite. *Astrophys. J.* **383**, 95–103 (1991).
- Jones, C. & Forman, W. The structure of clusters of galaxies observed with Einstein. *Astrophys. J.* **276**, 38–55 (1984).
- David, L. P., Jones, C. & Forman, W. ROSAT PSPC observations of cool rich clusters. *Astrophys. J.* **473**, 692–706 (1996).
- Knight, P. A. & Ponman, T. J. The properties of the hot gas in galaxy groups and clusters from 1-D hydrodynamical simulations—I. Cosmological infall models. *Mon. Not. R. Astron. Soc.* **289**, 955–972 (1997).
- Evvard, A. E. The intracluster gas fraction in X-ray clusters: constraints on the clustered mass density. *Mon. Not. R. Astron. Soc.* **292**, 289–297 (1997).
- Fukazawa, Y. et al. ASCA measurements of silicon and iron abundances in the intracluster medium. *Publ. Astron. Soc. Jpn* **50**, 187–193 (1998).
- Cavaliere, A., Menci, N. & Tozzi, P. The luminosity-temperature relation for groups and clusters of galaxies. *Astrophys. J.* **484**, L21–L24 (1997).
- Metzler, C. A. & Evvard, A. E. Simulations of galaxy clusters with and without winds I.—the structure of clusters. *Astrophys. J.* (submitted); available as preprint astro-ph/9710324 at (<http://xxx.lanl.gov>) (1997).
- Watt, M. P. et al. The morphology and dark matter distribution of the Coma cluster of galaxies from X-ray observations. *Mon. Not. R. Astron. Soc.* **258**, 738–748 (1992).
- Copi, C. J., Schramm, D. N. & Turner, M. S. Big-Bang nucleosynthesis and the baryon density of the Universe. *Science* **267**, 192–199 (1995).
- Miralda-Escudé, J., Cen, R., Ostriker, J. P. & Rauch, M. The Lyman-alpha forest from gravitational collapse in the cold dark matter plus lambda model. *Astrophys. J.* **471**, 582–616 (1996).
- Madau, P., Pozzetti, L. & Dickinson, M. The star formation history of field galaxies. *Astrophys. J.* **498**, 106–116 (1998).
- Heckman, T. M., Armus, L. M. & Miley, G. K. On the nature and implications of starburst-driven galactic superwinds. *Astrophys. J. Suppl.* **74**, 833–868 (1990).
- Bower, R. G. Entropy-driven X-ray evolution of galaxy clusters. *Mon. Not. R. Astron. Soc.* **288**, 355–364 (1997).
- Mushotzky, R. F. & Scharf, C. A. The luminosity-temperature relation at $z = 0.4$ for clusters of galaxies. *Astrophys. J.* **482**, L13–L16 (1997).
- Arnaud, M. & Evvard, A. E. The $L_X - T$ relation and intracluster gas fractions of X-ray clusters. *Mon. Not. R. Astron. Soc.* (submitted); preprint available as astro-ph/9806353 at (<http://xxx.lanl.gov>) (1998).
- Cen, R. & Ostriker, J. P. Most of the ordinary matter in the Universe is in warm/hot gas. *Science* (submitted); preprint available as astro-ph/9806281 at (<http://xxx.lanl.gov>) (1998).
- Snowden, S. L., McCammon, D., Burrows, D. N. & Mendenhall, J. A. Analysis procedures for ROSAT XRT PSPC observations of extended objects and diffuse background. *Astrophys. J.* **424**, 714–728 (1994).
- Ponman, T. J., Bourner, P. D. J., Ebeling, H. & Böhringer, H. A ROSAT survey of Hickson's compact galaxy groups. *Mon. Not. R. Astron. Soc.* **283**, 690–708 (1996).
- Yamashita, K. in *Frontiers of X-ray Astronomy* (eds Tanaka, Y. & Koyama, K.) 475–480 (Universal Academy Press, Tokyo, 1992).
- Butcher, J. A. *Extragalactic X-ray Astronomy with Ginga*. Thesis, Univ. Leicester (1994).
- Mulchaey, J. S., Davis, D. S., Mushotzky, R. F. & Burstein, D. The intra-group medium in poor groups of galaxies. *Astrophys. J.* **456**, 80–97 (1996).
- Ebeling, H., Mendes de Oliveira, C. & White, D. A. A2572 and HCG94—galaxy clusters but not as we know them: an X-ray case study of optical misclassifications. *Mon. Not. R. Astron. Soc.* **277**, 1006–1032 (1995).
- Raymond, J. C. & Smith, B. W. Soft X-ray spectrum of a hot plasma. *Astrophys. J. Suppl.* **35**, 419–439 (1977).
- Eke, V. R., Navarro, J. F. & Frenk, C. S. The evolution of X-ray clusters in low density universes. *Astrophys. J.* **503**, 569–592 (1998).

Acknowledgements. We thank R. Bower, S. White and P. Willmore for discussions. Data analysis was performed on the Starlink node at Birmingham. J.F.N. acknowledges the hospitality of the Max Planck Institut für Astrophysik in Garching during the preparation of this manuscript.

Correspondence and requests for materials should be addressed to T.J.P. (e-mail: tjp@star.sr.bham.ac.uk).

Detection of environmental fine structure in the low-energy β -decay spectrum of ^{187}Re

F. Gatti, F. Fontanelli, M. Galeazzi, A. M. Swift & S. Vitale

INFN Genoa and Genoa University, Via Dodecaneso 33, I-16146 Genoa, Italy

Determining whether neutrinos have mass is an important test of grand unified theories. Neutrino masses can in principle be measured from small distortions in the spectra of electrons emitted in low-energy β -decay processes. Detailed knowledge of the β -particles' spectral shape and the detector response is therefore required in order to distinguish^{1–5} real effects from instrumental artefacts. The interaction between the emitted β -particle and its local environment is predicted⁶ to produce oscillations in the β -particle spectrum, known as β environmental fine structure; the effect is analogous to extended X-ray absorption fine structure⁷ (EXAFS), which provides the basis for spectroscopic surface studies of local molecular structure. But the low energy resolution and high operating temperatures of traditional radiation detectors have precluded observation of β environmental fine structure. Cryogenic microcalorimeters, operated as particle detectors, offer a means of overcoming these problems, as they can reach energy resolutions up to ten times higher than traditional detectors⁸. Here we report the detection of β environmental fine structure in the β -decay spectrum of ^{187}Re , using a cryogenic microcalorimeter. Our results, which are in good agreement with recent theoretical predictions⁹, may facilitate studies of molecular or crystalline structures in a manner similar to EXAFS.

In X-ray absorption experiments, oscillatory fine structure modulation is observed in the photoabsorption cross-section due to the emitted photoelectron undergoing oscillations within the crystal lattice potential. If, in β -emission processes, the decaying atom is embedded in a molecule or a crystal, the ejected electron can be reflected from neighbouring atoms to produce a 'standing wave' in the region surrounding the emitting atom. This gives rise to oscillations (β environmental fine structure, hereafter, BEFS) in the energy spectrum of the emitted electron, whose period depends on the interatomic distance and whose amplitude depends on the electron-atom scattering cross-section.

The theoretical shape of the fractional BEFS modulation in β -decay, $\chi(E)$, can be calculated in a fashion analogous to that used in EXAFS^{7,10,11}, and is given by:

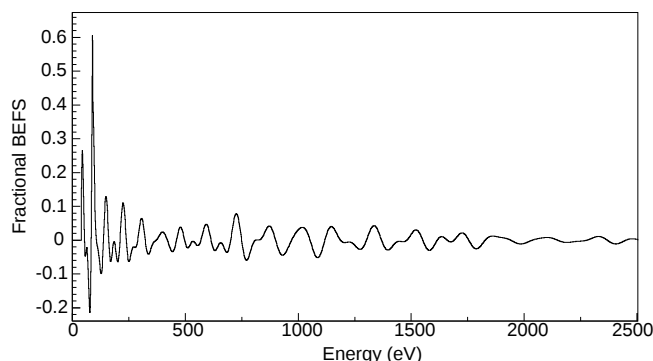


Figure 1 Theoretical BEFS shape for a rhenium single crystal at a temperature of 100 mK.

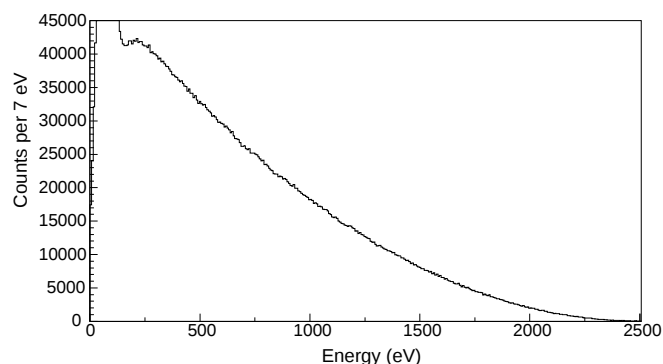


Figure 2 ^{187}Re β -particle spectrum obtained with our cryogenic microcalorimeter.

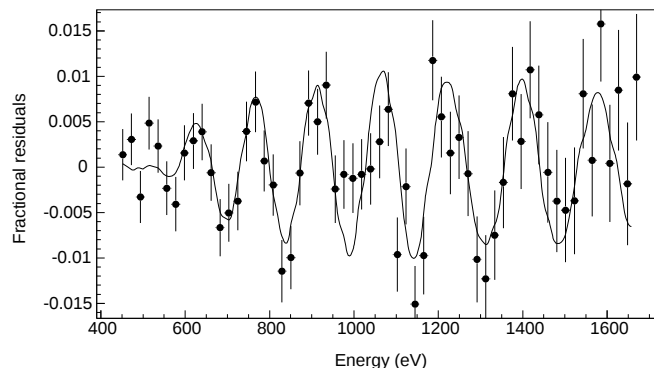


Figure 3 Experimental fraction residuals of fit to β -particle spectrum (filled circles), and convolution of theoretical BEFS calculation with the detector response (continuous line).

$$\chi(E) = \sum_i \frac{N_i |f_i(\pi)|}{kR_i^2} \sin(2kR_i + \Phi_i + 2\delta_0) e^{-\gamma R_i} e^{-2\sigma_i^2 k^2} \quad (1)$$

where the sum is over all neighbouring shells of atoms (labelled by i), $f_i(\pi)$ is the amplitude for an electron of energy $E = \hbar^2 k^2 / 2m$ to backscatter from atom i , $\Phi_i = \arg f_i(\pi)$, δ_0 is the s -wave phase shift induced by the decaying central atom, γ is the energy-dependent inverse β -electron mean free path, R_i is the distance between the central decaying atom and the surrounding shell of atoms, σ_i is the correction for zero-point and thermal lattice motion occurring in the Debye–Waller exponential term, N_i is the number of atoms in shell i , k is the wavevector, and m is the electron mass.

We have estimated the amplitude of BEFS using equation (1) for a rhenium single crystal at a temperature of 100 mK; we used the expression for Re atom shells from ref. 12. Rhenium forms a hexagonally close packed (h.c.p.) crystal; at room temperature, its lattice parameters are $a = 2.7609 \text{ \AA}$ and $c = 4.4576 \text{ \AA}$, whose ratio $c/a = 1.614$ is close to the ideal ratio for h.c.p. crystals, that is, $c/a = 1.6333$. There is a weak temperature dependence of the lattice parameters which extrapolates linearly to temperature $T = 0.01 \text{ K}$ to give $a = 2.7578 \text{ \AA}$, $c = 4.4514 \text{ \AA}$. The estimated BEFS is shown in Fig. 1. We note that the period of the oscillations ranges from $<100 \text{ eV}$ to $\sim 150 \text{ eV}$; a detector with good energy resolution is required to resolve the structure of the individual oscillations.

Until now, this effect has not been observed experimentally in any existing β -decay spectral study owing to the limitations of traditional radiation detectors. But cryogenic particle detectors bypass these limitations and are the best candidates for BEFS detection: they are characterized by a good energy resolution (down to $^{13}\text{7 eV}$ full-width at half-maximum, FWHM) and an extremely low energy threshold. In addition, the low operating temperature results in negligible thermal motion of the lattice which would otherwise reduce the coherence. The BEFS effect should dominate in the energy range less than $\sim 2 \text{ keV}$, because at high energies and correspondingly shorter wavelengths the effects of zero-point and thermal motion (the Debye–Waller effect) become more critical in the reduction of overall coherence.

The operating principles of cryogenic microcalorimeters as particle detectors have already been reported in detail by many authors⁸. A microcalorimeter is composed of an absorber (in which the electromagnetic energy is converted into phonons) with a strong thermal link to a phonon sensor; the phonon sensor is generally a thermistor whose resistance depends strongly on temperature. The microcalorimeter is weakly thermally linked to a refrigerator with a working temperature which ranges between 10 and 100 mK. The energy of the β -decay is first converted into thermal phonons by the

absorber, then into an electrical signal by the sensor. We note that this operating principle allows a wide choice of absorber material according to the goal of the experiment.

Our microcalorimeter consists of a 1.5-mg rhenium single crystal (abundance of ^{187}Re , 62.60%) as absorber, linked (using epoxy) to a neutron transmutation doped (NTD) germanium thermistor ($100 \times 100 \times 230 \text{ }\mu\text{m}$) as a sensor¹⁴. The detector is installed in a dilution refrigerator at an operating temperature of $\sim 60 \text{ mK}$. Mechanical, thermal and electrical connections between the microcalorimeter and the refrigerator are provided by two ultrasonically bonded $15\text{-}\mu\text{m}$ aluminium wires. The energy calibration is provided by a removable external X-ray source using ^{55}Fe (5,898-eV and 6,490-eV X-rays). Although the amount of energy is very small (down to $\sim 200 \text{ eV}$), the increase in temperature is not negligible (in the range of microkelvin) due to the extremely small heat capacity of the microcalorimeter at the working temperature. The complete ^{187}Re β -spectrum is shown in Fig. 2; the energy resolution of the detector is 80 eV FWHM. Due to this energy resolution we expect to detect oscillations of the order of 1–1.5% in the energy range 450–1,650 eV: at lower energies the oscillation period is smaller than the energy resolution, so the oscillations are strongly attenuated; at high energies, the oscillations are expected to decrease, becoming negligible in the end-point region.

The measured spectrum shown in Fig. 2 has been fitted by the theoretical spectral shape^{15,16} without BEFS; the difference between the real data and the fit in the range 450–1,650 eV is shown in Fig. 3, which also shows the BEFS theoretical calculation, convoluted with the detector response (continuous line). An oscillatory modulation may be seen in the data; a χ^2 statistical analysis indicates that the data are in disagreement with a straight line (no BEFS) with a 5σ confidence limit ($\chi^2 = 128.8$ with $n = 59$ degrees of freedom). When the BEFS is introduced, the experimental data are in agreement with the theoretical prediction within a 1σ confidence limit ($\chi^2 = 66$, $n = 59$).

Our experiment confirms the hypothesis of Koonin⁶ about the existence of BEFS, which should therefore be taken into account when analysing low-energy neutrino experiments with solid targets. Moreover, we have shown that cryogenic microcalorimeters are well suited to BEFS studies, due to their low operating temperature, good energy resolution and low energy threshold. Our results with cryogenic microcalorimeters should allow the use of BEFS as a kind of ‘internal EXAFS’ for structural studies; the relatively small size of the experimental set-up, the possibility of studying different absorber materials (crystals or molecules), the circumvention of the need for a large synchrotron light source and the present sophistication of EXAFS will probably encourage such studies. $\square$

Received 5 October; accepted 15 October 1998.

1. Simpson, J. J. Evidence of heavy neutrino emission in beta-decay. *Phys. Rev. Lett.* **54**, 1891–1893 (1985).
2. Sur, B. *et al.* Evidence for the emission of a 17-keV neutrino in the β decay of ^{14}C . *Phys. Rev. Lett.* **66**, 2444–2447 (1991).
3. Datar, V. M., Baba, C. V. K., Bhattacharjee, S. K., Bhuinya, C. R. & Roy, A. Search for a heavy neutrino in β -decay of ^{35}S . *Nature* **318**, 547–548 (1985).
4. Stoeffl, W. & Decman, D. J. Anomalous structure in the beta decay of gaseous molecular tritium. *Phys. Rev. Lett.* **75**, 3237–3240 (1995).
5. Belevsev, A. I. *et al.* Results of the Troitsk experiment on the search for the electron antineutrino rest mass in tritium beta decay. *Phys. Lett. B* **350**, 263–272 (1995).
6. Koonin, S. E. Environmental fine structure in low-energy β -particle spectra. *Nature* **354**, 468–469 (1991).
7. Ashley, C. A. & Doniach, S. Theory of extended x-ray absorption edge fine structure (EXAFS) in crystalline solids. *Phys. Rev. B* **11**, 1279–1288 (1975).
8. Booth, N., Cabrera, B. & Fiorini, E. Low temperature particle detectors. *Annu. Rev. Nucl. Part. Sci.* **46**, 471–532 (1996).
9. Swift, A. M. *The Study of the β Environmental Fine Structure Applied to the β -decay of ^{187}Re* 1–16 (Preprint INFN/BE-97/02, INFN, Genoa, 1997).
10. Lee, P. A., Citrin, P. H., Eisenberg, P. & Kincaid, B. M. Extended x-ray absorption fine structure—its strengths and limitations as a structural tool. *Rev. Mod. Phys.* **53**, 769–806 (1981).
11. Lee, P. A. & Pendry, J. Theory of the extended x-ray absorption fine structure. *Phys. Rev. B* **11**, 2795–2811 (1975).
12. Eckerlin, P. & Kandler, H. *Landolt Börnstein New Series* 6, 21 (eds Hellwege, K. H. & Hellwege, A. M.) (Springer, Berlin, 1971).
13. Stahle, C. K., Kelley, R. L., McCammon, D., Moseley, S. H. & Szymkowiak, A. E. Microcalorimeter array for high resolution soft x-ray spectroscopy. *Nucl. Instrum. Methods Phys. Res. A* **370**, 173–176 (1996).
14. Galeazzi, M. Status of the rhenium beta experiment for neutrino mass study. *Nucl. Phys. B (Proc. Suppl.)* **66**, 203–206 (1998).
15. Fontanelli, F., Galeazzi, M., Gatti, F., Swift, A. M. & Vitale, S. *Nucl. Instrum. Methods Phys. Res. A* (submitted).
16. Caso, C. *et al.* Review of particle physics. *Eur. Phys. J. C* **3**, 1–784 (1998).

Acknowledgements. This work was supported by INFN and the EC-HCM Programme 'Cryogenic Detectors'.

Correspondence and requests for materials should be addressed to S.V. (e-mail: vitale@ge.infn.it).

Lateral drag of spin coherence in gallium arsenide

J. M. Kikkawa & D. D. Awschalom

Department of Physics, University of California, Santa Barbara, California 93106, USA

The importance of spin-transport phenomena in condensed-matter physics has increased over the past decade with the advent of metallic giant-magnetoresistive systems and spin-valve transistors¹. An extension of such phenomena to semiconductors should create possibilities for seamless integration of 'spin electronics' with existing solid-state devices, and may someday enable quantum computing schemes using electronic spins as non-local mediators of coherent nuclear spin interactions². But to realize such goals, spin transport must be effected without destroying the relevant spin information. Here we report time-resolved optical studies of non-local Faraday rotation in n-type bulk gallium arsenide, which show macroscopic lateral transport of coherently precessing electronic spins over distances exceeding 100 micrometres. The ability to drag these spin packets by their negative charge, without a substantial increase in spin decoherence, is a consequence of the rather weak entanglement of spin coherence with orbital motion in this system³.

In Faraday-rotation measurements of spin precession within semiconductors, excitation by normally incident pulses of circularly polarized light produces a conduction-band spin imbalance in a superposition of spin states whose degeneracy is removed by a transverse (in-plane) magnetic field. The quantum coherent beating between these states results in Larmor precession of the classical spin vector, whose ensemble-average projection along the sample normal is measured by a time-delayed, linearly polarized probe pulse. The decay time of the transverse spin polarization, T_2^* , measures an upper limit on extra-electronic spin decoherence and conceals only rotationally invariant intra-electronic processes, which preserve the total electron magnetization⁴. The divergence

of T_2^* with reduced excitation and magnetic field complicates its measurement by traditional scans of the pump-probe delay, Δt , because relative polarization changes over nanosecond intervals are small in this regime and can be further obscured by slow Larmor precession. Hence, a method of resonant spin amplification is adopted in which the Larmor precession is driven into resonance with the optical excitation repetition interval, t_{rep} , by adjusting the applied field. The magnetic field width of the observed resonances is then used to accurately determine the transverse spin lifetime³.

Here we apply these methods to the study of mesoscopic lateral spin transport, fabricating samples that permit an externally applied, in-plane voltage. We study undoped and Si-doped ($1 \times 10^{16} \text{ cm}^{-3}$) GaAs substrates, each mechanically polished to a thickness of 30 μm and then uniformly contacted on one face with standard AuGeNi except in a $\sim 560\text{-}\mu\text{m}$ linear gap where optical studies are performed. A second sample of the Si-doped GaAs is prepared by wet etching to a thickness of $\sim 2.2 \mu\text{m}$, as determined by an increase in resistivity. Voltage leads are connected, and the samples are mounted in a magneto-optical cryostat for study in the Voigt configuration, the thicker samples freely suspended and the thinner sample bonded with transparent wax to a fused silica substrate. The gap orientation induces electric fields along the direction of the magnetic field. A miniature Hall sensor records the applied field, and the temperature is regulated at 1.6 K within superfluid ^4He . A mode-locked Ti:sapphire laser supplies a 1.51-eV pulse train at 76 MHz and 360- μW average power that is split into time-delayed pump and probe beams which overlap as they pass through the sample. The subsequent Faraday rotation of the probe beam is detected by a polarization bridge technique³. A stepper motor is used to adjust the lateral pump-probe separation on the

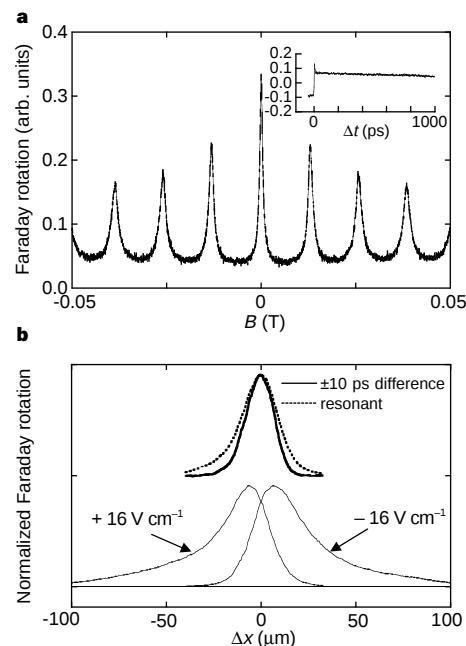


Figure 1 Pump-probe Faraday rotation measures spin polarization in time, magnetic field and space. **a**, Faraday rotation versus magnetic field for the 30- μm -thick, $1 \times 10^{16} \text{ cm}^{-3}$ Si-doped GaAs sample. Each peak results from the constructive reinforcement of successive pump pulses. Data are taken with a 100- μm pump diameter and a tightly focused probe. Inset, Faraday rotation versus pump-probe delay taken at a field of 0.0030 T. **b**, Spatial scans of the pump-probe interaction with 12- μm FWHM beams, normalized and offset in amplitude for clarity. The dotted line is the resonant response at $\Delta t = 0 \text{ ps}$ and $B = 0$, and the thick solid line shows the unbroadened spin injection profile, obtained by subtracting the corresponding data at $\Delta t = -10 \text{ ps}$. The lower data are taken at $B = 0$ with applied fields of $\pm 16 \text{ V cm}^{-1}$.

sample, Δx , along the direction of the in-plane electric field. In this manner, we are able to study the spatiotemporal profile of the optically excited spin polarization.

Figure 1a shows seven spin resonances near zero field in the Si-doped sample, obtained by scanning the magnetic field (B) at $\Delta t = 50$ ps. The inset shows the associated temporal behaviour of the spin polarization taken slightly off the $B = 0$ T resonance, with the offset at zero time delay arising from the injection of new spins on the arrival of the pump pulse. We note that pump–probe pairs reappear at 76 MHz, and past spin injections leave an imprint of negative polarization at $\Delta t < 0$ (ref. 3). The time scans are susceptible to inadvertent changes in beam overlap and focus with delay movement, so we reduce these effects by defocusing the pump. In contrast, spatial scans of the pump–probe overlap are taken at a fixed delay with the pump maximally focused, as shown in Fig. 1b for a field of $B = 0$ T. No precession occurs at this field, so spin accumulates from consecutive pump pulses, and the spatial profile at any given delay is broadened because of spin diffusion. Nonetheless, the spin injection profile may be extracted by taking the difference signal between scans obtained immediately before and after spin injection ($\Delta t = \pm 10$ ps). The upper part of Fig. 1b compares the spatial profile obtained by this method (solid line) to the wider profile taken at $\Delta t = -10$ ps (dotted line). The full-

width at half-maximum (FWHM) of the former ($18 \mu\text{m}$) sets the spatial resolution of our measurements.

The application of an in-plane electric field results in a macroscopic displacement of the electron spin polarization, as shown in lower portion of Fig. 1b for a field of $\pm 16 \text{ V cm}^{-1}$. We note that because these data are taken at $\Delta t = -10$ ps, the last pump pulse has arrived ~ 13 ns before the probe's arrival, so it is possible for there to be a greatly reduced signal at $\Delta t = 0$. In addition to the lateral displacement of the spin distribution, its profile becomes asymmetrical. Assuming that the total spin response $\mathbf{M}$ is the sum of non-interacting spin packets³, $\mathbf{M}(\Delta x, \Delta t) = \sum_n \mathbf{m}_n(\Delta x, \Delta t) = \sum_n \mathbf{m}_n(\Delta x, \Delta t + nt_{\text{rep}})$, we can understand this asymmetry as arising from a separation of the zero-field spin resonance into its constituents $\mathbf{m}_n$ created by distinct pump events. In the presence of an electric field, these spins no longer constructively reinforce each other at the resonance magnetic fields because they drift variable distances that are proportional to their ages. Despite an exponential time decay of the spin polarization that leads to a spatial decay along the drift direction, spin transport is observable at distances exceeding $100 \mu\text{m}$ in rather modest electric fields and mobilities³. Our data confirm that (1) the spin polarization measured is that of free electrons with a drift distance linear in electric field, and (2) the spins are carried by negative charges—that

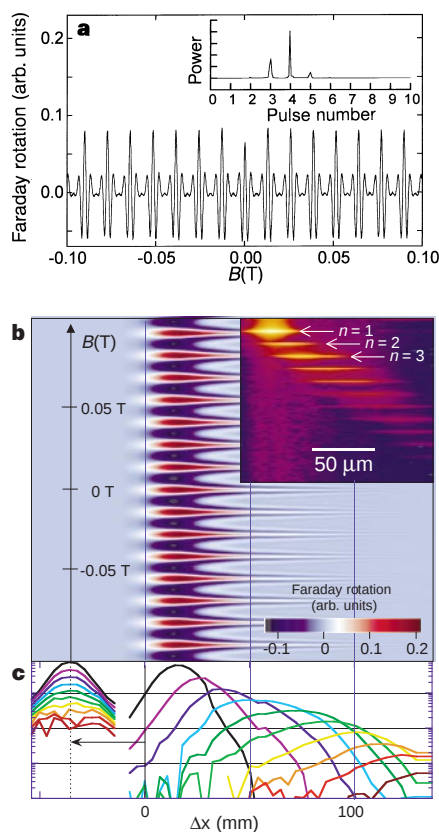


Figure 2 Macroscopic lateral spin transport as resolved by Fourier decomposition. **a**, Faraday rotation versus magnetic field at a displacement of $54 \mu\text{m}$ and an electric field of -37 V cm^{-1} . Inset, relative spectral powers of the $\mathbf{m}_n$ pulses at this location, indexed by n . **b**, Faraday rotation versus magnetic field and displacement. The field range is the same as in **a** and Δx values are taken from the axis of **c**. Inset, the spectral power of the $\mathbf{m}_n$ pulses over the range $-11 \mu\text{m} < \Delta x < 137 \mu\text{m}$. The first three pulses are labelled by their index, n . **c**, Amplitudes of the first ten $\mathbf{m}_n$ constituents versus displacement, obtained by explicit fits to the data in **b** and shown on a logarithmic scale. The $\mathbf{m}_n$ profiles at zero bias are also shown, displaced $-36 \mu\text{m}$ for clarity. Sample and excitation remain unchanged from Fig. 1b.

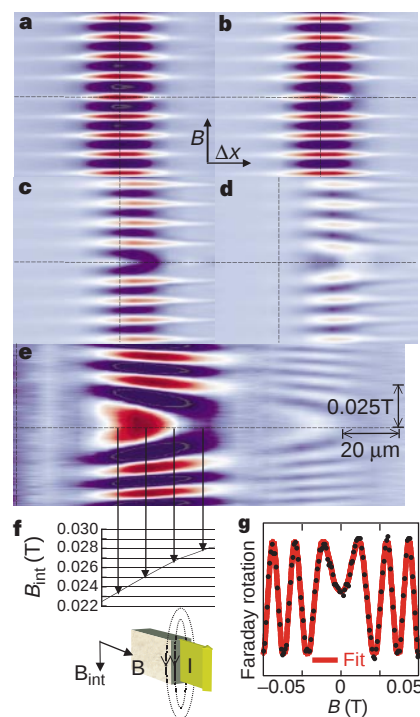


Figure 3 Self-induced dephasing of spin currents. **a–e**, Faraday rotation in the thinner sample versus magnetic field and displacement for voltages of 0, -9.2 , -18 , -37 and -92.5 V cm^{-1} , respectively. The data are taken at $\Delta t = -10$ ps, so that the youngest spin packet is ~ 13 ns old. The colour scheme is the same as in Fig. 2, except in **e** where the contrast has been enhanced. Vertical and horizontal dashed lines indicate zero displacement and field, respectively, and the scale bars in **e** apply to **a–d** as well. **f**, The internal magnetic field at positions $32.5 \mu\text{m} < \Delta x < 81.2 \mu\text{m}$ within the most recent spin injection, deduced from a model in which B_{int} is generated by currents within the sample (see drawing). The Δx scale is the same as in **e**. **g**, A cross-section of **e** at $\Delta x = 36 \mu\text{m}$. Points indicate the measured Faraday rotation versus magnetic field, and the solid red line is a fit to the model described within the text. Excitation remains unchanged from Fig. 1b.

is, we observe electron spins rather than hole spins. The latter has been assumed in a number of spin precession measurements on zinc blende semiconductors^{3–5}, and we believe that these data provide direct evidence that Faraday-rotation measurements are truly selective in measuring electron spins. We do not observe any spin polarization travelling oppositely to the electron spins, so the assumption that hole spins scatter rapidly in these systems appears to be valid^{5,6}. In the insulating (undoped) control sample, no changes in spin profile are observed at fields up to 400 V cm^{–1}.

To obtain a more precise measurement of spin drift in these systems, we note that each constituent of the spin resonance has a different periodicity, $g\mu_B(\Delta t + nt_{\text{rep}})/h$, in the applied magnetic field, B . Thus, by varying B at each displacement, we can separately identify the spatial extent of the various $\mathbf{m}_n$ by Fourier decomposition. Figure 2a shows the Faraday rotation versus magnetic field profile taken at a displacement of $\Delta x = 54 \mu\text{m}$ from the injection point and a pump–probe delay of $\Delta t = -10$ ps. An applied electric field of -37 V cm^{-1} creates a measurable spin polarization at this lateral position, and harmonic analysis (Fig. 2a inset) reveals that the oscillatory behaviour arises from Larmor precession of the third, fourth and fifth most recent pump pulses. By obtaining field scans over a range in Δx , we may track the spatial position of successive spin packets, as indexed by their injection time. Figure 2b shows a two-dimensional assembly of field scans similar to Fig. 2a, taken from $\Delta x = -54 \mu\text{m}$ to $+137 \mu\text{m}$. A narrowing of the spin resonances and an increasing periodicity in field accompanies lateral displacement, reflecting increased pulse ages with increased Δx . Figure 2b inset shows the corresponding harmonic power versus position, showing that pulse ages increase in steps of t_{rep} . The data can thus be used to mark the positions of individual spin packets, and indicate that spin drift is linear in time and corresponds to a drift mobility $\mu_d = 3 \times 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

By explicitly fitting the oscillatory data to obtain the amplitude of each spin packet at every position, we compare the profiles of the ten most recent spin injections at zero and -37 V cm^{-1} in Fig. 2c, where the former are laterally displaced for clarity. These data provide a quantitative measure of spin drift and diffusion over a 130-ns interval following spin injection. To remove artefacts from a strongly field-dependent transverse spin lifetime in the zero-bias sample, these data are obtained by fitting only the $B = 0$ T spin resonance. As seen on the logarithmic amplitude scale, the zero-bias spins decay exponentially with a characteristic time $T_2^* = 29$ ns. By fitting the broadening of the spin packet to $\sqrt{D_s(t + t_0)}$, where $t = \Delta t + nt_{\text{rep}}$ is the age of the n th pulse and t_0 adjusts for its initial width, we obtain a spin diffusion constant D_s that exceeds the electron diffusion constant, $D_e = \mu_d kT/e$, by more than one order of magnitude. This discrepancy suggests that spin diffusion here involves both electron and pure spin diffusion. Applying a voltage does not introduce any severe decoherence ($T_2^* \sim 17$ ns) and produces a more dramatic profile distortion during transport.

A more subtle effect is a suppression of the $B = 0$ T spin motion in this sample. Though difficult to distinguish within the colour scale of Fig. 2b, a relative dip in spin polarization appears for $\Delta x \neq 0$ and fields below a few gauss, as seen in Fig. 2a. Figure 3a–e shows that these effects become stronger in the thinner sample where, as the electric field is increased from 0 to -92.5 V cm^{-1} , the $B = 0$ T resonance disappears completely and is eventually replaced by neighbouring resonances. These data suggest that electric current flow generates an additional internal magnetic field, B_{int} , that dominates near zero external magnetic field. Figure 3g shows a cross-section of the most recent spin packet, separated from other spins by an electric field of -92.5 V cm^{-1} . The data deviate from an expected cosinusoidal behaviour and are more accurately described by an effective field strength, $B_{\text{eff}}^2 = B_{\text{int}}^2 + B^2$, with $B_{\text{int}} = 0.023$ T. A change in the effective magnetic field within the profile of each packet is evident from the shift of the resonances to lower field

strengths at larger displacements. This behaviour cannot be explained by a variation of g -factors throughout the profile of the spin packet because the Larmor frequency at larger magnetic fields is independent of Δx (not shown). Hence by fitting the field dependence of the Faraday rotation at different Δx , we obtain a profile of the internal magnetic field strength throughout the most recent pulse (Fig. 3f, top). This field distribution leads to inhomogeneous dephasing of the spin direction within each spin packet and scales linearly with the electric field. Similar fits for the thicker sample indicate an internal field that is two orders of magnitude smaller.

Although the magnetic field generated by electric current is of the correct geometry (Fig. 3f, bottom), it changes sign in passing from the front to the back face of the sample whereas the data are well fitted by a single internal field at each Δx . Moreover, for uniform current flow the field strength should scale linearly with the thickness of the sample, while the opposite trend is observed in our data; rough estimates of the surface field strength give the correct order of magnitude for the thicker sample but fail for the thinner sample. One possible explanation is that the optical injection creates a significantly non-uniform, and perhaps filamentary, current flow⁷. Further consideration must also be given to the inherent spin splitting that is known to occur in the conduction band of GaAs away from zero wavevector⁸. We find it difficult to identify B_{int} solely with the corresponding effective field, however, as the latter is linearly proportional to and uniquely determined by the drift velocity⁸ whereas the former is dramatically different in thick and thin samples with the same drift velocity. □

Received 4 August; accepted 5 November 1998.

1. Prinz, G. Spin-polarized transport. *Phys. Today* **48**, 58–63 (1995).
2. Kane, B. E. A silicon based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
3. Kikkawa, J. M. & Awschalom, D. D. Resonant spin amplification in GaAs. *Phys. Rev. Lett.* **80**, 4313–4316 (1998).
4. Kikkawa, J. M., Smorchkova, I. P., Samarth, N. & Awschalom, D. D. Room-temperature spin memory in two-dimensional electron gases. *Science* **277**, 1284–1287 (1997).
5. Worsley, R. E., Traynor, N. J., Grevatt, T. & Harley, R. T. Transient linear birefringence in GaAs quantum wells. *Phys. Rev. Lett.* **76**, 3224–3227 (1996).
6. Oestreich, M. *et al.* Temperature and density dependence of the electron Landé g -factor in semiconductors. *Phys. Rev. B* **53**, 7911–7916 (1996).
7. Niedernostheide, F.-J., Hirsinger, J., Prettl, W., Novak, V. & Kostial, H. Oscillations of current filaments in n-GaAs caused by a magnetic field. *Phys. Rev. B* **58**, 4454–4458 (1998).
8. Kalevich, V. K. & Korenov, V. L. Effect of electric field on the optical orientation of 2D electrons. *Pis'ma Zh. Eksp. Teor. Fiz.* **52**, 230–235 (1990) [*JETP Lett.* **52**, 230–235 (1990)].

Acknowledgements. We thank S. J. Allen for discussions and acknowledge support from the ARO and NSF QUEST STC.

Correspondence and requests for materials should be addressed to D.D.A. (e-mail: awsch@physics.ucsb.edu).

Entropic trapping of macromolecules by mesoscopic periodic voids in a polymer hydrogel

Lei Liu^{†,†}, Pusheng Li^{†,†} & Sanford A. Asher^{*}

^{*} Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

The separation of macromolecules such as polymers and DNA by means of electrophoresis, gel permeation chromatography or filtration exploits size-dependent differences in the time it takes for the molecules to migrate through a random porous network. Transport through the gel matrices, which usually consist of full swollen crosslinked polymers^{1–11}, depends on the relative size of

[†] Present address: Praxair Service Technologies Inc., 1555 Main St, Indianapolis, Indiana 46254, USA (L.L.); Epsilon Inc., HPOS, 50 Cambridge St, Burlington, Massachusetts 01803, USA (P.L.).

the macromolecule compared with the pore radius. Sufficiently small molecules are thought to adopt an approximately spherical conformation when diffusing through the gel matrix¹, whereas larger ones are forced to migrate in a snake-like fashion^{3–5}. Molecules of intermediate size, however, can get temporarily trapped in the largest pores of the matrix, where the molecule can extend and thus maximize its conformational entropy. This ‘entropic trapping’ is thought to increase the dependence of diffusion rate on molecular size^{6–16}. Here we report the direct experimental verification of this phenomenon. Bragg diffraction from a hydrogel containing a periodic array of monodisperse water voids confirms that polymers of different weights partition between the hydrogel matrix and the water voids according to the predictions of the entropic trapping theory. Our approach might also lead to the design of improved separation media based on entropic trapping.

Figure 1 illustrates the fabrication of a polymerized crystalline colloidal array (CCA) of spherical water voids. An aqueous suspension of ~100-nm-diameter monodisperse silica spheres was induced to self-assemble into a three-dimensional CCA by removing the ionic impurities from the medium with ion-exchange resin. This self-assembly, driven by interparticle electrostatic repulsion due to negative particle surface charges, minimizes the total system energy^{17–24}. The array adopts either a body-centred cubic or face-centred cubic structure^{17–24}. Similar to atomic crystals diffracting X-rays, CCAs Bragg-diffract light in the near-infrared to near-ultraviolet spectral regions, depending on the interparticle spacing and the system refractive index^{17,21–24}.

We immobilized the CCA in a polyacrylamide hydrogen network^{21–24}, by mixing acrylamide, *N,N'*-methylene-bisacrylamide, and photo-initiator 2,2-diethoxyacetophenone into the silica CCA dispersion. Photo-polymerization results in a ~100-μm-thick hydrogel film, where the silica spheres are locked in their cubic lattice. These polymerized CCA hydrogel films contain a high polymer and crosslinker content, and are sufficiently robust to allow further chemical modification. This silica polymerized CCA was used as a template to create a polymerized CCA of ~100-nm-diameter spherical water voids (HPCCA) by hydrofluoric acid etching. Both diffraction and gravimetry demonstrate complete removal of the silica colloids with little effect on the hydrogel network.

We soaked the HPCCA film in deionized water solutions of sodium polystyrene sulphonate (NaPSS) of different molecular masses with narrow molecular-mass distributions ($M_w/M_n \approx 1.1$, where M_w and M_n are the mass-average and number-average molecular masses) for 3–10 days until equilibrium was established. Figure 2 shows the experimental method and the measured HPCCA diffraction spectra before and after soaking in NaPSS solution.

We model the CCA Bragg diffraction by combining dynamical diffraction and light scattering theory^{23–26}, and analyse the diffracted

wavelength and intensity changes to sensitively probe the chemical composition in the voids and in the gel medium. Both the diffraction intensity (I_D) and wavelength (λ_D) depend upon the refractive indices^{23–26}:

$$-\log\left(\frac{I_T}{I_0}\right) = -\log\left(1 - \frac{I_D}{I_0}\right) = f(n_{\text{void}}, n_{\text{gel}}, n_{\text{crystal}}) \quad (1)$$

$$\lambda_D = g(n_{\text{crystal}}, n_{\text{gel}}) \quad (2)$$

where I_T and I_0 are the transmitted and incident light intensities. The refractive indices are related to the chemical compositions by:

$$n_{\text{void}} = n_{\text{water}}(1 - C_H) + n_{\text{PSS}}C_H \quad (3)$$

$$n_{\text{gel}} = n_{\text{AMD}}\phi_a + (1 - \phi_a)[n_{\text{water}}(1 - C_G) + n_{\text{PSS}}C_G] \quad (4)$$

$$n_{\text{crystal}} = n_{\text{gel}}(1 - \phi_{\text{void}}) + n_{\text{void}}\phi_{\text{void}} \quad (5)$$

where n_{water} , n_{AMD} , and n_{PSS} represent the refractive indices of water, polyacrylamide and NaPSS, respectively, ϕ_{void} is the volume fraction of the voids in the whole system, ϕ_a is the volume fraction of the polyacrylamide network in the gel medium, C_H and C_G are the NaPSS concentrations (in wt%) in the voids and in the gel medium. C_G here is normalized with respect to only the total solution volume of the interstitial gel medium; the polyacrylamide chain network is excluded.

More uptake of NaPSS in the voids than in the gel medium decreases the refractive-index difference between the voids and the gel medium, which decreases the diffraction intensity. Uptake of NaPSS also increases the overall crystal refractive index, and slightly red-shifts the diffracted wavelength.

Figure 3a shows the dependence of the measured NaPSS partition coefficient $K_{\text{HG}} (=C_H/C_G)$ on the NaPSS molecular mass. The natural logarithm of K_{HG} initially increases almost linearly with NaPSS molecular mass and then levels off. For the highest molecular masses, C_H is ~4 times larger than that in the gel medium.

The observed dependence of K_{HG} on the molecular mass of NaPSS must result from entropic effects because enthalpic interactions between the NaPSS chains and the hydrogel network should show little molecular mass dependence. The embedded voids possess essentially the same chemical properties as the rest of the gel medium except that the diameter of the voids (~100 nm) is much larger than the mean pore diameter of the gel network (pore diameter is in the nanometre size range; ref. 27). Large flexible NaPSS chains preferentially partition into the voids in order to maximize their conformational entropy.

When the configuration of a flexible chain polymer is treated using self-avoiding random-walk statistics, the constrained conformational entropy in a box of a particular shape can be calculated^{13,14,28–30}. The entropy dominated partition coefficient K_j

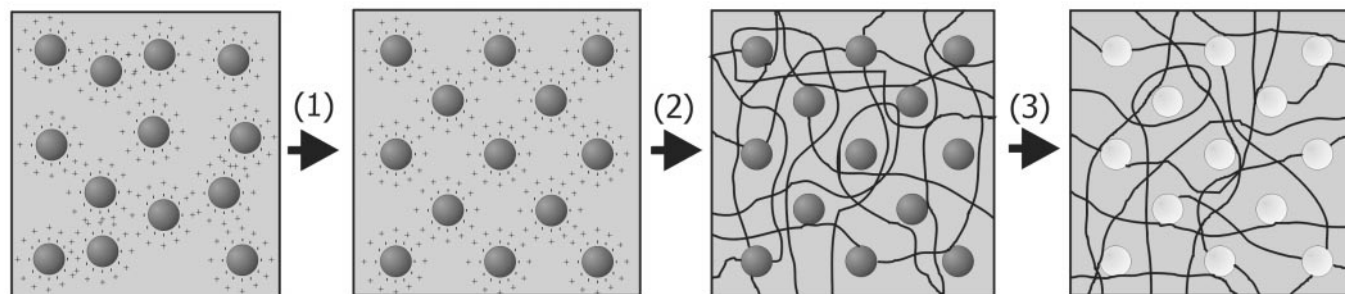


Figure 1 Fabrication of hydrogel with a cubic array of spherical water voids. (1), Monodisperse silica spheres electrostatically self-assemble into a crystalline colloidal array (CCA) in aqueous solution. (2), The silica CCA is immobilized within

a random polyacrylamide hydrogel network matrix (a polymerized CCA). (3), Silica spheres are etched out with hydrofluoric acid to result in a polymerized CCA of water voids (HPCCA).

between two confining cavities i and j will scale as:

$$-\ln K_{ij} = -\ln \left(\frac{C_i}{C_j} \right) \approx N \left[\left(\frac{1}{R_i} \right)^{\frac{1}{v}} - \left(\frac{1}{R_j} \right)^{\frac{1}{v}} \right] \quad (6)$$

where C_i and C_j are the monomer concentrations in cavities i and j , respectively, N is the polymer molecular mass, v is the Flory universal exponent, and R_i and R_j are the characteristic radii of cavities i and j . This thermodynamic model assumes well-defined rigid boundary conditions, and is only valid for cavities sufficiently large to accommodate the entire polymer chain^{13,14,28}.

At low molecular mass, our results agree qualitatively with these predictions. When the pore sizes of both the voids and the gel network are much larger than the NaPSS molecular sizes, there is little constraint in either region; no partitioning occurs. As the molecular mass increases, $\ln K_{HG}$ increases linearly with molecular mass as predicted by entropic trapping theory.

But at the highest molecular masses, this linear relationship fails, presumably because the theory is based on impermeable rigid boundaries and the assumption that a single cavity contains the entire polymer chain. In the hydrogel matrix, the 'boundaries' of the 'cavities' are poorly defined because the aqueous solution forms a continuous phase interwoven in a three-dimensional fashion with the crosslinked polyacrylamide chains. When the NaPSS molecular size is much larger than the mean mesh spacing of the gel medium, the chain will simultaneously occupy two or more adjacent pores of that medium. Each section will behave like an independent polymer chain of smaller molecular mass. As predicted by Muthukumar *et al.*^{13,14}, this effect will diminish the

molecular-mass dependence of entropic trapping and the effective size of the hydrogel pores (R_j) will appear to increase with the NaPSS molecular mass (equation (6)).

Entropic trapping also depends on concentration. Figure 3b shows the dependence of K_{HG} on the NaPSS reservoir solution concentration (C_S) for NaPSS with the largest molecular mass (1.2×10^6 dalton, radius of gyration $R_g \approx 34$ nm) and the smallest (1.6×10^3 dalton). For the lowest molecular mass, K_{HG} slightly increases monotonically with C_S . In contrast, for the highest molecular mass, K_{HG} sharply increases until it reaches a maximum at $\sim 5\%$ C_S , and then decreases as C_S increases further.

The observed concentration dependence of K_{HG} may result from NaPSS intermolecular interactions such as exclusion and entanglement². When polymer chains enter a cavity, they will encounter intermolecular exclusion from volumes already occupied by other polymer chains. The effective cavity volume will decrease as polymer chains enter. As exclusion is more effective in the hydrogel pores than in the voids, K_{HG} should in general increase with concentration. The more rapid initial increase of K_{HG} with C_S seen for the higher molecular mass NaPSS chains may be due to the fact that polymer chains of higher molecular mass possess larger hydrodynamic volumes than those of lower mass. At identical concentrations, larger polymer chains thus exclude each other more effectively than smaller chains.

However, an increase in K_{HG} causes an increasing concentration difference between the voids and the gel medium, thus exaggerating volume exclusion effects in the voids at higher NaPSS concentrations. In the case of the highest molecular mass NaPSS, the solution concentration in the voids may be sufficiently high to cause entanglement, while the gel medium remains in the dilute solution regime. As pointed out by Guillot *et al.*⁶, both effects counteract entropic trapping from the hydrogel network and could eventually result in a decrease of K_{HG} with C_S .

We expect that the fabrication approach we have used in this study could serve as a general motif to design new separation materials in which macromolecules could be retarded, trapped and separated. Such materials might also be useful as specific biochemical trapping materials for applications in drug delivery

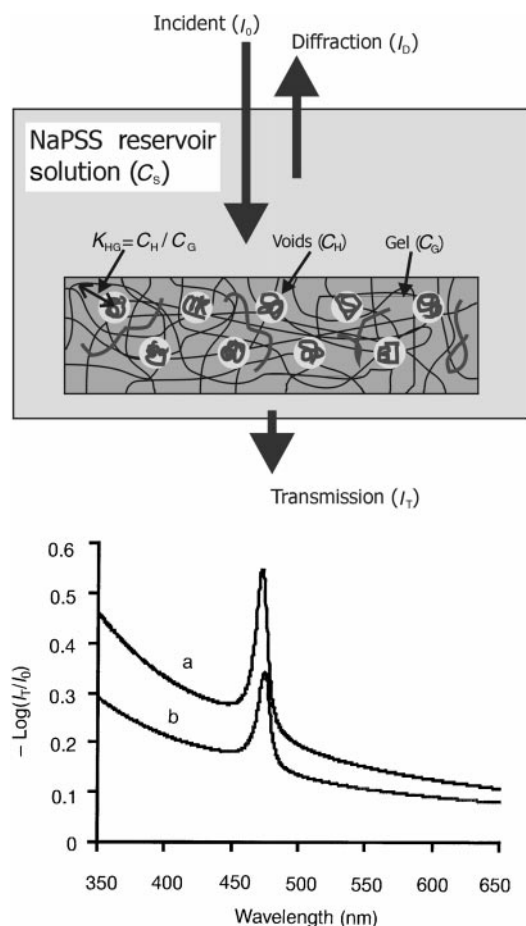


Figure 2 Bragg diffraction probes the chemical composition of HPCCA. Top, Experimental setup. Bottom, measured Bragg diffraction spectra from HPCCA film before (trace a) and after (trace b) exchange with a NaPSS reservoir solution.

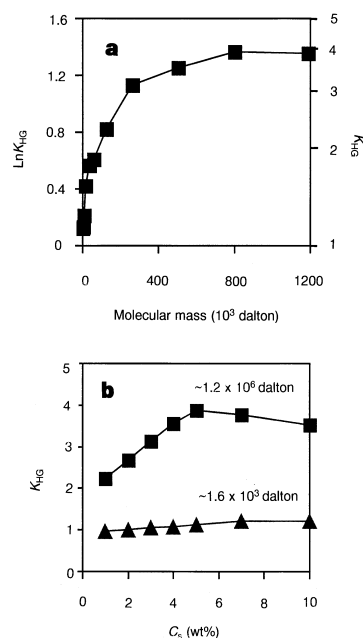


Figure 3 Dependence of NaPSS partitioning on molecular mass and concentration. Dependence of the NaPSS partition coefficient, $K_{HG} (=C_H/C_G)$, between the voids and the gel medium on: **a**, the molecular mass at 5 wt% reservoir concentration, and **b**, the reservoir concentration (C_S). The standard deviation in the partition coefficient is likely to be significantly smaller than the size of the data points.

and controlled-release processes, or as semi-homogeneous micro-reactors for applications in organic, bioengineering and combinatorial synthesis. In these applications, Bragg diffraction from the arrays could be used to probe the existence and location of chemical species as a means to monitor physicochemical processes in real time. □

Received 23 April; accepted 11 September 1998.

- Rodbard, D. & Chrambach, A. Unified theory for gel electrophoresis and gel filtration. *Proc. Natl. Acad. Sci. USA* **65**, 970–977 (1970).
- Teraoka, I. Interferometric study of transition from weak to strong penetration of a polymer solution into a porous silica bead. *Macromolecules* **29**, 2430–2439 (1996).
- Lerman, L. S. & Frisch, H. L. Why does the electrophoretic mobility of DNA in gels vary with the length of the molecule? *Biopolymers* **21**, 995–997 (1982).
- Lumpkin, O. J., Dejdard, P. & Zimm, B. H. Theory of gel electrophoresis of DNA. *Biopolymers* **24**, 1573–1593 (1985).
- Slater, G. W. & Noolandi, J. On the reptation theory of gel electrophoresis. *Biopolymers* **25**, 431–454 (1986).
- Guillot, G., Leger, L. & Rondelez, F. Diffusion of large flexible polymer chains through model porous membranes. *Macromolecules* **18**, 2531–2537 (1985).
- Rotstein, N. A. & Lodge, T. P. Tracer diffusion of linear polystyrenes in poly(vinyl methyl ether) gels. *Macromolecules* **25**, 1316–1325 (1992).
- Smisek, D. L. & Hoagland, D. A. Electrophoresis of flexible macromolecules: evidence for a new mode of transport in gels. *Science* **248**, 1221–1223 (1990).
- Muthukumar, M. & Hoagland, D. A. Evidence for entropic barrier transport of linear, star, and ring macromolecules in electrophoresis gels. *Macromolecules* **25**, 6696–6698 (1992).
- Rousseau, J., Drouin, G. & Slater, G. W. Entropic trapping of DNA during gel electrophoresis: effect of field intensity and gel concentration. *Phys. Rev. Lett.* **79**, 1945–1948 (1997).
- Noolandi, J., Rousseau, J. & Slater, G. W. Self-trapping and anomalous dispersion of DNA in electrophoresis. *Phys. Rev. Lett.* **58**, 2428–2431 (1987).
- Slater, G. W. & Wu, S. Y. Reptation, entropic trapping, percolation, and Rouse dynamics of polymers in “random” environments. *Phys. Rev. Lett.* **75**, 164–167 (1995).
- Baumgärtner, A. & Muthukumar, M. A trapped polymer chain in random porous media. *J. Chem. Phys.* **87**, 3082–3088 (1987).
- Muthukumar, M. & Baumgärtner, A. Effects of entropic barriers on polymer dynamics. *Macromolecules* **22**, 1937–1941 (1989).
- Kim, H., Chang, T., Yohanan, J. M., Wang, L. & Yu, H. Polymer diffusion in linear matrices: polystyrene in toluene. *Macromolecules* **19**, 2737–2744 (1986).
- Nemoto, N., Kishine, M., Inoue, T. & Osaki, K. Tracer diffusion of linear polystyrene in entanglement networks. *Macromolecules* **23**, 659–664 (1990).
- Hiltner, P. A. & Krieger, I. M. Diffraction of light by ordered suspensions. *J. Phys. Chem.* **73**, 2386–2389 (1969).
- Clark, N. A., Hurd, A. J. & Ackerson, B. J. Single colloidal crystals. *Nature* **281**, 57–60 (1979).
- Kesavamoorthy, R., Tandon, S., Xu, S., Jagannathan, S. & Asher, S. A. Self-assembly and ordering of electrostatically stabilized silica suspensions. *J. Colloid Interface Sci.* **153**, 188–198 (1992).
- Asher, S. A. Crystalline narrow band radiation filter. US Patent Nos 4,627,689 and 4,632,517 (1986).
- Weissman, J. M., Sunkara, H. B., Tse, A. S. & Asher, S. A. Thermally switchable periodicities and diffraction from mesoscopically ordered materials. *Science* **274**, 959–960 (1996).
- Holtz, J. H. & Asher, S. A. Polymerized colloidal crystal hydrogel films as intelligent chemical sensing materials. *Nature* **389**, 829–832 (1997).
- Rundquist, P. A., Photinos, P., Jagannathan, S. & Asher, S. A. Dynamical Bragg diffraction from crystalline colloidal arrays. *J. Chem. Phys.* **91**, 4932–4941 (1989).
- Liu, L., Li, P. & Asher, S. a. Fortuitously superimposed lattice plane secondary diffraction from crystalline colloidal arrays. *J. Am. Chem. Soc.* **119**, 2729–2732 (1997).
- Zachariasen, W. H. *Theory of X-ray Diffraction in Crystals* (Dover, New York, 1957).
- Van de Hulst, H. C. *Light Scattering by Small Particles* (Dover, New York, 1957).
- Righetti, P. G. Macroporous gels: facts and misfacts. *J. Chromatogr. A* **698**, 3–17 (1995).
- Casassa, E. F. Equilibrium distribution of flexible polymer chains between a macroscopic solution phase and small voids. *Polymer Lett.* **5**, 773–778 (1967).
- Flory, P. J. *Principles of Polymer Chemistry* (Ithaca, New York, 1953).
- De Gennes, P. G. *Scaling Concepts in Polymer Physics* (Ithaca, New York, 1979).

Acknowledgements. We thank M. D. Morris, D. Pine, J. Holtz, J. M. Weissman and G. Pan for discussions. This work was supported by the Office of Naval Research and the US National Science Foundation.

Correspondence and requests for materials should be addressed to S.A.A. (e-mail: asher+@pitt.edu).

A nanomechanical device based on the B–Z transition of DNA

Chengde Mao, Weiqiong Sun, Zhiyong Shen & Nadrian C. Seeman

Department of Chemistry, New York University, New York, New York 10003, USA

The assembly of synthetic, controllable molecular mechanical systems^{1–7} is one of the goals of nanotechnology. Protein-based molecular machines, often driven by an energy source such as ATP, are abundant in biology^{8,9}. It has been shown previously that branched motifs of DNA can provide components for the assembly of nanoscale objects¹⁰, links¹¹ and arrays¹². Here we show that

such structures can also provide the basis for dynamic assemblies: switchable molecular machines. We have constructed a supramolecular device consisting of two rigid DNA ‘double-crossover’ (DX) molecules connected by 4.5 double-helical turns. One domain of each DX molecule is attached to the connecting helix. To effect switchable motion in this assembly, we use the transition between the B and Z^{13,14} forms of DNA. In conditions that favour B-DNA, the two unconnected domains of the DX molecules lie on the same side of the central helix. In Z-DNA-promoting conditions, however, these domains switch to opposite sides of the helix. This relative repositioning is detected by means of fluorescence resonance energy transfer spectroscopy, which measures the relative proximity of two dye molecules attached to the free ends of the DX molecules. The switching event induces atomic displacements of 20–60 Å.

It has been recognized for a long time that the B–Z transition could be used to drive a DNA-based mechanical device¹⁵, but the development of such a device required the discovery of a rigid DNA motif to provide a fixed level arm: the rigidity of the DX molecule¹⁶ is the key element that enables the construction of this device.

A molecular model of the system is shown at the top of Fig. 1. A long central helix is flanked on either end by two ‘DAO’ molecules, where DAO (for double-crossover, antiparallel, odd-number) means that this is a DNA double-crossover molecule whose helices are antiparallel to each other, and whose crossover points are separated by an odd number (three) of double helical half-turns¹⁷. The ends of all helices have been closed with dT₄ hairpin loops (made up of four deoxythymidines); consequently, the entire construct consists of three cyclic strands of DNA, the two blue strands on the ends, and the red strand that contains a yellow central

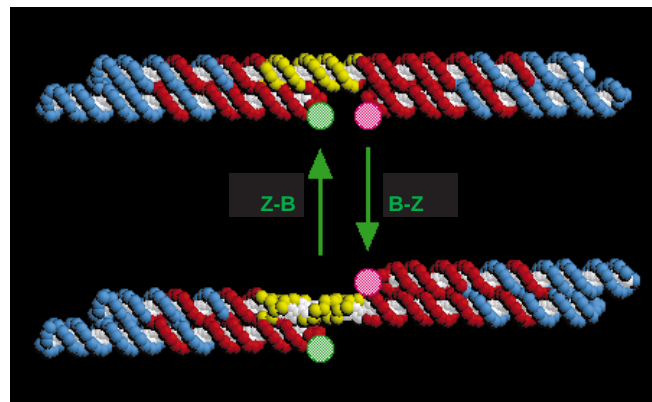


Figure 1 Design of the nanomechanical device. The device consists of two DNA double crossover (DX) molecules with the DAO motif¹⁷ connected by 4.5 turns of DNA between the nearest crossover points. Top, molecular model of the molecule constructed entirely from right-handed B-DNA. Each nucleotide is shown as two spheres, a coloured one for the backbone and a white one for the base. Three cyclic strands are shown, one in the centre drawn as a red strand with a central yellow segment, and two blue strands on the ends that are each triply catenated to the red strand. Fluorescent dyes are drawn schematically as stippled green (fluorescein) and magenta (Cy3) circles attached to the free hairpins near the middle of the molecule. At the centre of the connecting helix is a 20-nucleotide region of proto-Z DNA in the B-DNA conformation, shown in yellow. When the B–Z transition takes place, this same yellow portion becomes left-handed Z-DNA (bottom). When the transition occurs, the two DX molecules change their relative positions, increasing the separation of the dyes. It is possible to cycle this system in both directions.

segment. The outer strands are each linked three times to the central strand. The double helix connecting the two DAO molecules contains the base-paired sequence $d(CG)_{10}$, a 'proto-Z' sequence that can be converted to left-handed Z-DNA at high ionic strength; this is the portion of the red strand drawn in yellow. The top part of Fig. 1 shows this yellow region in the right-handed form, B-DNA, and the bottom part shows it in the left-handed form, Z-DNA. The change in twist from B-DNA to Z-DNA is estimated to be about -128° for each $d(CG)$ dinucleotide¹³. Thus, ignoring the B-Z junctions, whose twist is unknown¹⁸, the change in twist between the B and the Z forms of this molecule is expected to be about -3.5 turns, resulting in a transition roughly like that depicted in Fig. 1. Overall, the transition consists largely of a rotary motion; however, the separation of the base pairs along the helix axis ($3.7 \text{ \AA}$) is slightly longer in Z-DNA than in B-DNA ($3.4 \text{ \AA}$) (ref. 13), which is expected to result in a $6\text{-\AA}$ expansion (visible in Fig. 1) in the length of the central helix axis.

Two fluorescent dyes—a donor, fluorescein, and an acceptor, the sulphoindocarbocyanine, Cy3 (ref. 19)—are represented in Fig. 1 as green and magenta stippled circles, respectively. The drawing shows that the distance between them is expected to change when the transition occurs. The signal in fluorescence resonance energy transfer (FRET) is proportional to the inverse sixth power of the separation of the dyes, providing a means to monitor the transition. The experiment is predicated on the assumption that well-defined distances will characterize the separation of the dyes in the two states. Well-defined distances can exist only if the parts of the molecule maintain their structural integrity; initial attempts to demonstrate motion in devices containing three-arm DNA

branched junctions instead of DX molecules were unsuccessful, owing to the flexibility²⁰ of those components. This problem has been solved, because the DX motif behaves as a rigid unit¹⁶.

The device was constructed by ligating a series of oligonucleotides, using methods described previously¹⁰. The molecule was isolated from denaturing gels, and its successful construction was demonstrated by restricting the individual strands to yield target linear, circular and catenated products (data not shown). To account for the uncertainty involving the twist of the B-Z junction, we have made similar molecules containing $d(CG)_9$ and $d(CG)_{11}$; moreover, we have varied the separation of the inner crossover points in steps of two nucleotides, over the range 45–49. Of the nine molecules tested, the lowest relative energy transfer under Z-promoting conditions (compared with a non-proto-Z control) was obtained from the molecule whose separation of 47 nucleotides contained $d(CG)_{10}$. To confirm that the B-Z transition can occur in this molecule, we have prepared the central 47-nucleotide double helix, and have shown that a mixed B-Z circular dichroism spectrum is obtained in 5M NaCl (data not shown), unlike a control lacking the proto-Z sequence.

Figure 2 illustrates the results of taking this system through successive cycles of B-promoting and Z-promoting conditions. The molecule used for these experiments contains $d(m^5CG)_{10}$ (in which the 5-position of cytosine is methylated, a modification that increases the propensity of DNA to undergo the B-Z transition²¹; as a result, we are able to promote the transition with a 10 mM cacodylate buffer (pH 7.5) containing 0.25 mM $Co(NH_3)_6Cl_3$, 100 mM $MgCl_2$ and 100 mM NaCl, whereas B-DNA is present when the same buffer contains 10 mM $MgCl_2$, 100 mM NaCl and no $Co(NH_3)_6Cl_3$. Figure 1 shows that the larger separation is designed to occur in the Z-form, and that the response to Z-promoting conditions should be to lengthen the distance and decrease the FRET signal between the two dyes. This is seen to happen in Fig. 2, both when energy transfer is measured by decrease of donor fluorescence or by increase of acceptor fluorescence. Figure 2 also illustrates the results obtained with a control molecule, one designed to be incapable of undergoing the B-Z transition. In this case, energy transfer appears to be at roughly the same level throughout the experiments, a level appropriate only to the B-state of the molecule.

The experiments reported above show that it is possible to construct a nanomechanical device from DNA DX molecules. The absolute distances derived from energy-transfer measurements are 7.0 nm for B conditions and 8.9 nm for Z conditions, somewhat longer than those derived from model building²², 5.1 nm and 7.0 nm; nevertheless, the ratio of measured distances, 1.27, is similar to the model-building ratio of 1.37. It should be possible to incorporate this mechanical control in any figure or array produced by DNA nanotechnology²³, so long as a free swivel containing proto-Z DNA can be included in the design. In addition, it should be possible to change the relative positions of proteins or other large molecules connected to DX units²⁴, in the same way that the relative positions of fluorescent dyes have been changed here. This capability should enable the study of proximity effects in chemical and biological systems. It is difficult to predict whether this system could also be used to power a nanoscale motor; we have shown that the B-Z transition can provide two equilibrium structures in different positions, but the ability to transmit force depends on the structural and dynamic features of the transition itself.

We have recently reported a system in which double-stranded DNA branch migration, characterized by a screw motion, is controlled by the addition of ethidium to the solution²⁵. However, that system involves a long piece of circular DNA whose geometry is less well-defined than the molecule reported here. The synthetic systems mentioned above produce molecular motions of the order of 10 Å. By contrast, the atoms in the rotated DNA helical domain experience displacements that range from 20 to 60 Å. The fusion of those

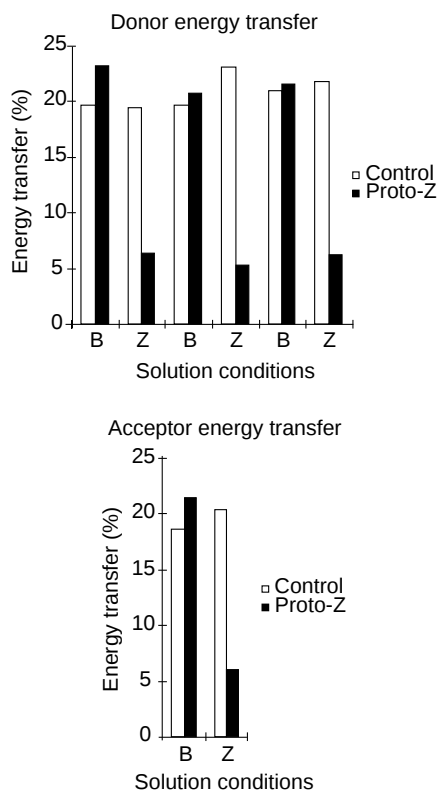


Figure 2 FRET demonstration of nanomechanical motion. Top, energy transfer measured from the decrease in donor fluorescence when cycling the system alternately through B and Z conditions. A similar result is obtained when the increase in acceptor fluorescence is used (bottom). In both graphs, the molecule containing the proto-Z segment is contrasted with a control that lacks it. Only the molecule containing the proto-Z segment shows a change in energy transfer in Z-promoting conditions.

differently triggered smaller systems with the synthetic DNA system reported here could provide a variety of complex nanomechanical motions. □

Methods

Molecular sequences. The sequences of the blue strands (Fig. 1) are: 5'-TCGAGCCTGTACCTGCCAAGTCAGACTGTGGACTTCGCATCACGGACGACTCATCTTTTGATGAGTCGTGGTACAGGCTCGAGTTGCTTTTGCAAC-3' and 5'-AATTCAGTGAACGACCGCATTTGGCTGACACCATACGCCGGAATCCTGACCATCTGTTTTCAGATGGTCACCGTTACACTGGAATTCGCACTTTGTGCG-3'. The sequence of the red strand is 5'-TGCCGATCCTTT(Cy3)TTGGATCGGCAAGCTTCTGTGACCCAATGCGGTGGATTCCGGCGTATGGACGCGAGTC*GC*GC*GC*GC*GC*GC*GC*GAAGCCGCTCTGGCTCGACCACGCGGTTGGATCCTTFTTGATCCAACGGCTGGACAGTCTGACTGGCACCCTGATGCGAAGTCCTGGTCAGAGCCAGCGGCTTC*GC*GC*GC*GC*GC*GC*GC*GA-CTGCGTGGAAAGCT-3', where Cy3 is attached to amino-modifier-C6-dT (Glen Research), F is the product of incorporating fluorescein-ON phosphoramidite (Clontech), and C* is 5-methyl-cytidine. The control molecule replaces the proto-Z sequence (yellow segment) with TGGACACTAAGCTATTCATC and its complement.

Calculation of energy transfer. Energy transfer is calculated following ref. 19. Energy transfer involving the donor, ET(D), is derived from $ET(D) = [F_{490}^{517}(D) - F_{490}^{517}(DA)]/F_{490}^{517}(D)$, where $F_{490}^{517}(D)$ is the fluorescence of the donor at its wavelengths of maximum emission (517 nm) and excitation (490 nm) in the molecule containing only the donor and $F_{490}^{517}(DA)$ is the same quantity in the doubly labelled molecule. Energy transfer involving the acceptor is derived from $ET(A) = (\epsilon^{550}(DA)/\epsilon^{490}(DA)) \cdot [(ratio)_A - \{F_{490}^{565}(A)/F_{550}^{565}(A)\}]$, where $(ratio)_A = [F_{490}^{565}(DA) - \epsilon F_{490}^{565}(D)]/F_{550}^{565}(DA)$; $\epsilon^{550}(DA)$ and $\epsilon^{490}(DA)$ are the extinction coefficients at 550 and 490 nm, respectively, of the doubly labelled molecule (their ratio is 3.65 in B-DNA conditions and 3.0 in Z-DNA conditions), $F_{490}^{565}(A)$ is the fluorescence at 565 nm, the maximum emission wavelength, of the acceptor when excited at 490 nm, $F_{550}^{565}(A)$ is the fluorescence at 565 of the acceptor when excited at 550 nm, its maximum excitation wavelength; the quantities in $(ratio)_A$ are similarly defined for the donor molecule (D) or the doubly labelled molecule (DA), and ϵ is a normalization constant used to equalize donor emission in the (D) and (DA) molecules at 517 nm. The refractive indices of the B-DNA and Z-DNA promoting buffers are 1.338 and 1.339, respectively. The anisotropies of the dyes, measured in singly labelled molecules, were found to be 0.136 for fluorescein and 0.259 for Cy3 in both buffers, similar to values reported earlier¹⁹, and validating the use of 2/3 for the value of K^2 , the scalar orientation factor for dipole-dipole coupling²⁶. R_0 (the characteristic transfer distance) is calculated to be 55.7 Å (B-DNA conditions) or 56.1 Å (Z-DNA conditions), in agreement with previous work¹⁹. In experiments involving $Co(NH_3)_6Cl_3$ the fluorescein fluorescence is quenched by 15%, but this appears to have no effect on the relative energy transfer or on the comparison between molecules with and without the proto-Z sequence.

Molecular modelling. The molecular model in Fig. 1 was drawn using the Insight II Molecular Modeling System, version 95.0.6 (Biosym Technologies). The coordinates for B-DNA and Z-DNA were taken from that program's data base. The dT₃ hairpin loops were sketched to resemble the dT₄ hairpin loops whose structure was determined in ref. 27.

Received 6 May; accepted 10 September 1998.

- Dado, G. & Gellman, S. H. Redox control of secondary structure in a designed peptide. *J. Am. Chem. Soc.* **115**, 12609–12610 (1993).
- Handel, T. M., Williams, S. A. & deGrado, W. F. Metal ion dependent modulation of the dynamics of a designed protein. *Science* **261**, 879–885 (1993).
- Bissell, R. A., Códova, E., Kalfer, A. E. & Stoddart, J. F. A chemically and electrochemically switchable molecular shuttle. *Nature* **369**, 133–137 (1994).
- Zelkovich, L., Libman, J. & Shanzler, A. Molecular redox switches based on chemical triggering of iron translocation in triple-stranded helical complexes. *Nature* **374**, 790–792 (1995).
- Livoreil, A. et al. Electrochemically and photochemically driven ring motions in a dissymmetrical copolymer [2] catenate. *J. Am. Chem. Soc.* **119**, 12114–12124 (1997).
- Murakami, M., Kawabuchi, A., Kotoo, K., Kumitake, M. & Nakashima, N. A light-driven molecular shuttle based on a rotaxane. *J. Am. Chem. Soc.* **119**, 7605–7606 (1997).
- Zahn, S. & Canary, J. W. Redox-switched exciton-coupled circular dichroism: A novel strategy for binary molecular switching. *Angew. Chem. Int. Edn* **37**, 305–307 (1998).
- Nedelec, F. J., Surrey, T., Maggs, A. C. & Leibler, S. Self-organization of microtubules and motors. *Nature* **389**, 305–308 (1997).
- Alberts, B. The cell as a collection of protein machines. *Cell* **92**, 291–294 (1998).
- Chen, J. & Seeman, N. C. The synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).

- Mao, C., Sun, W. & Seeman, N. C. Assembly of Borromean rings from DNA. *Nature* **386**, 137–138 (1997).
- Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
- Rich, A., Nordheim, A. & Wang, A. H.-J. The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53**, 791–846 (1984).
- Pohl, F. M. & Jovin, T. M. Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC). *J. Mol. Biol.* **67**, 375–396 (1972).
- Seeman, N. C. in *Biomolecular Materials* (ed. Viney, C., Case, S. T. & Waite, J. H.) 123–134 (Symp. Proc. Vol. 292, Materials Research Soc, 1993).
- Li, X., Yang, X., Qi, J. & Seeman, N. C. Antiparallel DNA double crossover molecules as components for nanoconstruction. *J. Am. Chem. Soc.* **118**, 6131–6140 (1996).
- Fu, T.-J. & Seeman, N. C. DNA double crossover structures. *Biochemistry* **32**, 3211–3220 (1993).
- Sheardy, R. D. et al. Sequence dependence of the free energy of B-Z junction formation in deoxypolynucleotides. *J. Mol. Biol.* **231**, 475–488 (1993).
- Jares-Erijman, E. A. & Jovin, T. M. Determination of DNA helical handedness by fluorescence resonance energy transfer. *J. Mol. Biol.* **257**, 597–617 (1996).
- Ma, R.-L., Kallenbach, N. R., Sheardy, R. D., Petrillo, M. L. & Seeman, N. C. 3-arm nucleic acid junctions are flexible. *Nucleic Acids Res.* **14**, 9745–9753 (1986).
- Behe, M. & Felsenfeld, G. Effects of methylation on a synthetic polynucleotide: The B-Z transition in poly (dG-m5dC).poly (dG-m5dC). *Proc. Natl Acad. Sci. USA* **78**, 1619–1623 (1981).
- Seeman, N. C. Physical models for exploring DNA topology. *J. Biol. Struct. Dyn.* **5**, 997–1004 (1988).
- Seeman, N. C. DNA nanotechnology: novel DNA constructions. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 225–248 (1998).
- Smith, S. S. et al. Nucleoprotein-based nanoscale assembly. *Proc. Natl Acad. Sci. USA* **94**, 2162–2167 (1997).
- Yang, X., Liu, B., Vologodskii, A. V., Kemper, B. & Seeman, N. C. Torsional control of double stranded DNA branch migration. *Biopolymers* **45**, 69–83 (1998).
- Clegg, R. M. et al. Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry* **31**, 4846–4856 (1992).
- Chattopadhyaya, R., Grzeskowiak, K. & Dickerson, R. E. Structure of a T₄ hairpin loop on a Z-DNA stem and comparison with A-RNA and B-DNA loops. *J. Mol. Biol.* **211**, 189–210 (1990).

Acknowledgements. We thank S. Zhang for work done on previous systems, and D. Millar, N. Geacintov and R. Sheardy for advice and for the use of equipment. This work was supported by the Office of Naval Research, the National Institute of General Medical Sciences, the National Science Foundation/DARPA and the Information Directorate of the Air Force Research Laboratory (Rome, NY).

Correspondence and requests for materials should be addressed to N.C.S. (e-mail: ned.seeman@nyu.edu).

Sulphidic Mediterranean surface waters during Pliocene sapropel formation

Hilde F. Passier*†, Hendrik-Jan Bosch*, Ivar A. Nijenhuis*, Lucas J. Lourens*, Michael E. Böttcher‡†, Anke Leenders*§, Jaap S. Sinninghe Damsté*§, Gert J. de Lange* & Jan W. de Leeuw*§

* Institute of Paleoenvironments and Paleoclimate Utrecht (IPPU), Geochemistry and Micropaleontology, Utrecht University, PO Box 80021, NL-3508 TA Utrecht, The Netherlands

† Institute for Chemistry and Biology of the Marine Environment (ICBM), Oldenburg University, Germany

§ Netherlands Institute for Sea Research (NIOZ), Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

Sapropels—organic-matter rich layers—are common in Neogene sediments of the eastern Mediterranean Sea. The formation of these layers has been attributed to climate-related increases in organic-matter production^{1–3} and increased organic-matter preservation due to oxygen depletion in more stagnant bottom waters^{2,3}. Here we report that eastern Mediterranean Pliocene sapropels⁴ contain molecular fossils of a compound (isorenieratene) known to be synthesized by photosynthetic green sulphur bacteria, suggesting that sulphidic (euxinic)—and therefore anoxic—conditions prevailed in the photic zone of the water column. These sapropels also have a high trace-metal content, which is probably due to the efficient scavenging of these metals by precipitating sulphides in a euxinic water column. The abundance and sulphur-isotope composition of pyrite are consistent with iron sulphide formation in the water column. We conclude that basin-wide water-column euxinia occurred over substantial

† Present addresses: Paleomagnetic Laboratory, Budapestlaan 17, NL-3584 CD Utrecht, The Netherlands (H.F.P.); Max-Planck-Institute for Marine Microbiology, Biogeochemistry, Celsiusstrasse 1, D-28359 Bremen, Germany (M.E.B.).

Table 1 Data from Pliocene sapropels

Sapropel code	ODP sapropel* (cm)	Age (Myr)	Water depth (m)	Depth (mbsf)†	Av. δ C _{dll} (p.p.m.)	Av. δ Moll (p.p.m.)	Av. δ NiII (p.p.m.)	Av. δ Sell (p.p.m.)	Av. δ VII (p.p.m.)	Av. δ C _{org} II (wt%)	AR# (g C m ⁻² yr ⁻¹)	Av. δ Ball (p.p.m.)	S _{pyr} (wt%)	δ^{34} S _{pyr} (‰)	Iso. deriv.
	Single sample‡ (cm)														
i-176	160-967C-6H-2, 30-44 36-37.7	1.808	2,553	48.4	n.d.¶	98 (32)	190 (2)	13 (121)	430 (5)	5.0 (37) 7.4	1.8	n.d.¶	5.8	-41.4	++
i-282A	160-964D-10H-1, 103-110 107.5-108	2.943	3,660	80.2	34 (130)	195 (16)	317 (9)	29 (131)	1,317 (13)	15.3 (149) 21.1	6.2	1,603 (13)	2.4	-36.1	+
i-282B	160-967C-8H-4, 114-130 122.5-123	2.943	2,553	72.2	3 (45)	202 (6)	183 (4)	22 (52)	555 (6)	8.4 (63) 14.5	3.4	1,029 (8)	2.8	-37.4	+
i-282C	160-969E-6H-6, 27-39 32-33	2.943	2,201	50.7	21 (270)	407 (26)	299 (4)	24 (17)	1,937 (19)	17.7 (81) 23.5	7.1	1,316 (11)	2.1	-42.0	++
i-290	160-964E-6H-5, 77-103 99-100	3.036	3,661	82.9	6 (61)	45 (6)	291 (9)	n.d.¶	304 (5)	5.0 (76) 23.6	2.0	n.d.¶	3.6	-39.5	+
i-292A	160-966C-5H-3,4, 148-13 10-11	3.058	926	42.6	1 (9)	133 (21)	101 (3)	n.d.¶	340 (4)	6.0 (24) 11.6	1.2	n.d.¶	1.1	-39.1	-
i-292B	160-967C-8H-6, 40-57 46-47	3.058	2,553	74.5	8 (51)	224 (18)	174 (4)	n.d.¶	918 (7)	13.2 (64) 20.3	5.3	n.d.¶	1.9	-37.3	+

* Stratigraphic position of the complete sapropel (ODP coding).

† Stratigraphic interval of the single sample of which C_{org}, S_{pyr}, δ^{34} S_{pyr} and iso. deriv. are listed.

‡ m.b.s.f., metres below sea floor.

§ Av., average.

¶ Enrichment factors with respect to average non-sapropel sediments around the sapropels are indicated in brackets.

¶ n.d., no data.

AR, accumulation rate of organic carbon = $\frac{1}{n} \sum (C_{org} \text{ (wt\%)} \times \text{sediment accumulation rate (cm kyr}^{-1}) \times \text{dry density (g cm}^{-3})/10) / n$. Here n is the number of samples in the sapropel.

☆ Iso. deriv., isorenieratene derivatives: ++, abundant; +, trace; -, not detected.

periods during Pliocene sapropel formation in the eastern Mediterranean Sea, and that the ultimate degradation of the increased organic-matter production was strongly influential in generating and sustaining the euxinic conditions.

Geochemical analyses were performed on seven Upper Pliocene sapropels obtained from several eastern Mediterranean ODP Leg 160 sites (Table 1; Fig. 1). Correlation of the sapropels with the insolation cycles of the Sun revealed that these sapropels formed at the same time as three sapropels in the Punta Piccola section (Sicily, Italy), i-292, i-290 and i-282, and sapropel i-176 of the Vrica and Singa sections (Calabria, Italy), which have ages of 3.058, 3.036, 2.943 and 1.808 Myr, respectively⁵. The sapropels are distinct, dark-grey to black, sediment layers with a thickness ranging from centimetres to decimetres, embedded in light-grey to brown hemipelagic carbonate oozes. Fine lamination of the sapropels indicates the absence of bioturbation and, therefore, bottom-water oxygen depletion⁴. The sapropels contain considerable amounts of isorenieratene derivatives, trace metals and pyrite (Table 1; Fig. 2).

The occurrence of isorenieratene derivatives in the sapropels

analysed (Table 1) indicates a periodic overlap of the photic and euxinic zones during sapropel formation. Isorenieratene is a carotenoid believed to be biosynthesized exclusively by the brown strain of photosynthetic green sulphur bacteria (Chlorobiaceae)⁶. As Chlorobiaceae are phototrophic and require free sulphide as an electron donor to fix CO₂, their occurrence is restricted to euxinic waters in the photic zone. These bacteria fix CO₂ in the reverse tricarboxylic acid cycle, resulting in a biomass anomalously enriched in ¹³C (ref. 7). The isorenieratene derivatives are indeed 13–14‰ enriched in ¹³C relative to algae-derived biomarkers, which confirms that they originate from green sulphur bacteria^{8–11}. The presence of ¹³C-enriched isorenieratene-derived compounds in the Pliocene sapropels therefore provides strong evidence that the lower photic zone contained sulphide over substantial periods of time during sapropel formation. This situation is comparable to the present-day euxinic Black Sea, where green sulphur bacteria occur at the lower photic zone¹² and isorenieratene derivatives are found in the sediment¹⁰.

Euxinic conditions in the water column during sapropel deposition

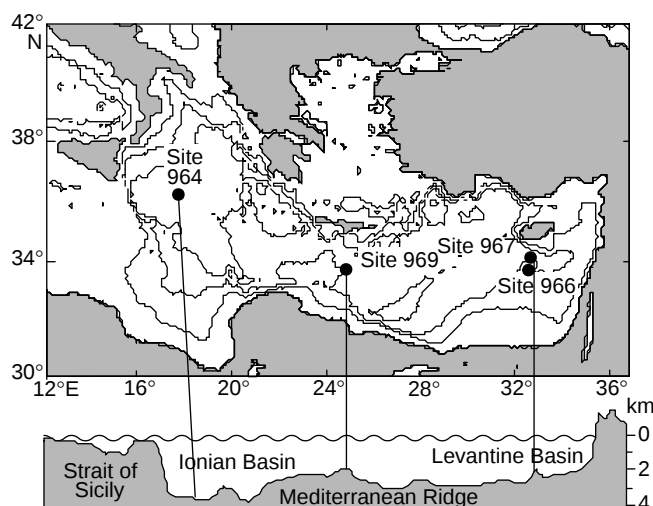


Figure 1 Locations of the drilling sites in the eastern Mediterranean.

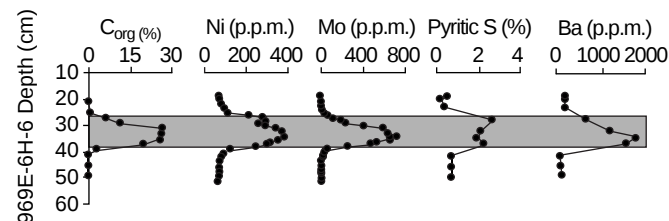


Figure 2 Sapropel compositional data. Content versus depth profiles of C_{org}, Ni, Mo, pyritic S, and Ba in and around ODP sapropel 160-969E-6H-6, 27-39 cm; the shaded area indicates the position of the sapropel.

explain the exceptionally high contents in the sapropels of trace metals (Table 1; Fig. 2) scavenged by organic matter and by metal sulphide precipitation. These trace-metal enrichments strongly exceed the maximum amount that can be supplied to the sapropels by plankton or diagenesis¹³. Therefore, scavenging by metal sulphide precipitation under euxinic conditions, which is also thought to have played an important role in transferring trace metals to similarly enriched Cretaceous black shales¹⁴, must have been operative during formation of the studied sapropels.

Pyrite (FeS₂) enrichments in the sapropels (Fig. 2; Table 1) are attributed to organic-matter decomposition via microbial sulphate reduction and subsequent reaction of the reduced sulphur with iron during sapropel formation^{15,16}. Supply of detrital iron and iron diffusion towards the sapropels cannot adequately explain the amount of pyrite found in the sapropels. The integrated pyritic iron content of sapropel i-282C is 6 mmol cm⁻², whereas the supply of detrital iron and upward iron diffusion from underlying sediments during sapropel formation is only 2 mmol cm⁻² (calculated at Site UM26¹⁷ near Site 969). A euxinic water column provides an additional iron source as a result of iron sulphide formation in the water column. Additional iron for this iron sulphide formation can be supplied after its liberation from oxides and hydroxides at the margins of a euxinic basin in response to a fluctuating oxycline, as occurs at present in the Black Sea^{18,19}. Isotopic compositions of pyritic sulphur ($\delta^{34}\text{S}_{\text{pyr}}$; defined in Methods) are consistent with a euxinic water column. Values of $\delta^{34}\text{S}_{\text{pyr}}$ in the sapropels are about -40‰ (Table 1), indicating a depletion by ~60‰ compared to seawater sulphate (present-day value is +20.6‰; ref. 15). To achieve fractionations of sulphur isotopes of this magnitude between sulphate and pyrite, sulphate reduction must proceed with a continuous supply of abundant dissolved sulphate (open system), and must be followed by further ³⁴S depletion in the sulphide pool by reoxidation and disproportionation processes in the sulphur cycle²⁰. In the present-day Black Sea^{20,21}, sulphide reoxidation occurs at the chemocline, which results in lighter isotopic compositions of sulphide in the euxinic water column. Directly below sapropels (Fig. 2), pyrite is enriched relative to normal organic-matter-poor marine sediments. This indicates that sulphide diffused into the organic-matter-poor sediment below as a result of iron limitation for pyrite formation within the sapropel^{15,16}.

The combined observations (presence of ¹³C-rich isorenieratene, enrichments of trace metals and ³²S-rich pyrite, and lamination of the sapropels) indicate that the eastern Mediterranean water column must have been euxinic from the bottom up into the lower photic zone over substantial periods during Pliocene sapropel formation. This is, to our knowledge, the first evidence for such extreme oxygen depletion during sapropel deposition. Both increased primary productivity (leading to a higher oxygen demand) and the development of more stagnant conditions (resulting in reduced oxygen replenishment of bottom waters) have been proposed to cause sapropel formation¹⁻³, and are probably related to the climate-induced^{22,23} increase in freshwater inputs from the African and European continents^{2,3,22}.

We now consider whether primary productivity increased during Pliocene sapropel formation. The euxinic eastern Mediterranean and the present-day euxinic Black Sea have similar depths, and the organic matter in the sapropels of both basins is predominantly of marine origin^{4,24}. Thus, applying the euxinic preservation efficiency observed in the modern deep Black Sea (at maximum, 5% of the primary produced organic matter²⁴) to the present-day average primary productivity in the open eastern Mediterranean (26 g C m⁻² yr⁻¹; ref. 25) we find a maximum accumulation rate of organic carbon (C_{org}) of 1 g C m⁻² yr⁻¹. In contrast, the accumulation rate of C_{org} in Pliocene sapropels (Table 1) calculated from the C_{org} contents and sediment properties⁴ is 4 g C m⁻² yr⁻¹ on average, illustrating that primary productivity must have increased significantly during Pliocene sapropel formation.

The enrichment of barium (Table 1; Fig. 2) in the sapropels also points strongly to a productivity increase. Sedimentary barium may serve as a palaeoproductivity indicator because it originates from barium uptake into barite in decaying organic matter associated with primary productivity in the surface waters²⁶⁻²⁸. Sapropel i-282C has an intermediate barium accumulation compared to its lateral equivalents i-282A and i-282B (Table 1; Fig. 1); the barium enrichment is therefore representative of the average deep basin, minimizing possible sediment-focusing effects. The barium profile (Fig. 2) is smooth, so no dissolution and remobilization of barite took place after burial, indicating that the barium enrichment is original. This is because porewater sulphate was never depleted, which is also indicated by $\delta^{34}\text{S}_{\text{pyr}}$ (ref. 16). Because the barium enrichment in i-282C is original and representative, barium contents can be used to calculate export palaeoproductivity (the part of the primary productivity of organic matter exported from the photic zone to deep waters) via an empirical formula²⁸; this gives an average export productivity of 46 g C m⁻² yr⁻¹ during sapropel deposition. Some limitations are associated with this calculation. The formula was determined for the oxic ocean²⁸, which may not be comparable to the eastern Mediterranean during sapropel deposition. However, the upper part of the photic zone, where barite predominantly forms²⁶⁻²⁸, remained oxic. Also, barite accumulation was not limited by sulphate, because the concentration of sulphate in sea water is so high that it was not significantly depleted by sulphate reduction in the water column, which is also indicated by $\delta^{34}\text{S}_{\text{pyr}}$ (ref. 16). Barite formation related to productivity has been shown to occur in euxinic basins²⁷ and in oxygen-minimum zones²⁹, as well as in the open ocean^{26,28}. Euxinic conditions may, however, lead to reduced formation and poorer preservation of barite²⁶⁻²⁸, which means that the calculated export productivity is a minimum value.

Could the extreme oxygen depletion in the water column result from the observed productivity increase? This oxygen depletion can be estimated from the average export productivity calculated from sedimentary barium (for example, ≥ 46 g C m⁻² yr⁻¹ for sapropel i-282C) and the amount of organic matter that accumulated in the sediment (7 g C m⁻² yr⁻¹ for sapropel i-282C), corrected for the part of the organic matter that was mineralized by microbial sulphate reduction. Sulphate reduction produces 1 mol of sulphide for every 2 mol of organic matter that are mineralized ($2\text{CH}_2\text{O} + \text{SO}_4^{2-} \rightarrow \text{H}_2\text{S} + 2\text{HCO}_3^-$). Any sulphide that escaped from burial was reoxidized at the chemocline, so part of the organic matter was indirectly mineralized by oxygen. The sulphide that was buried represents the part of organic matter that was truly mineralized by sulphate reduction; in sapropel i-282C, this is 20% of the originally accumulated C_{org}. The oxygen demand in the water column during sapropel deposition (calculated from carbon fluxes) is $\geq 1 \times 10^{13}$ mol O₂ yr⁻¹, assuming that 138 mol O₂ are required for 106 mol "CH₂O" (ref. 25) and that sapropels occur in 80% of the eastern Mediterranean (excluding the area shallower than 200 m). The estimated present-day oxygen supply to the eastern Mediterranean through deep-water formation in the Levantine and Adriatic basins is also 1×10^{13} mol O₂ yr⁻¹ (ref. 25). In the present-day eastern Mediterranean (where export production is 6 to 12 g C m⁻² yr⁻¹; ref. 25), the amount of oxygen used for organic matter decomposition is less than the amount supplied, resulting in an oxic water column²⁵. In contrast, our estimates indicate that the oxygen demand resulting from increased export production of organic matter during Pliocene sapropel formation is of the same order of magnitude as the present-day oxygen supply; this means that oxygen in the water column may have become depleted so that sulphate reduction led to euxinic conditions. The calculated oxygen demand is a good first-order estimate, based on the average deep eastern Mediterranean. But variations in the content, composition and burial efficiency of organic matter, especially across the basin margins, may affect the oxygen demand.

The above palaeoflux calculations suggest that water-column euxinia could have been achieved even when continued circulation supplied oxygen to deep waters. The trace-metal budget for the studied Pliocene sapropels strongly favours continued circulation during deposition, as this would have provided an extra input of trace metals, required to account for the observed amounts of trace metals in the sapropels. For example, the total amount of nickel sequestered in sapropel i-282 (average of sites 964, 967 and 969), assuming 80% coverage, is 8×10^{13} g. The water column in a fully stagnant eastern Mediterranean, however, contains only $\sim 2 \times 10^{12}$ g Ni ($\sim 3\%$)¹³, and rivers, assuming global average river composition and a present-day discharge, can supply only $\sim 8 \times 10^{12}$ g Ni (that is, about 10%) during sapropel formation. Unfortunately, exact river discharges of trace metals during sapropel formation are unknown. A ten-times increase in river water discharge during sapropel formation, however, is highly unrealistic^{2,3}. Consequently, it is unlikely that Ni discharge increased to such an extent. Budgets for other trace metals in sapropel i-282 and i-292 yield comparable results¹³. It therefore seems that stagnation of the water column was not as important as generally believed^{2,3}. □

Methods

Sediment samples (0.5–1 cm resolution) were freeze-dried and ground. Organic-carbon contents (C_{org}) were determined with an NCS-analyser after carbonate removal with 1M HCl (ref. 13). Sediments were dissolved in $HClO_4-HNO_3-HF$; dried residues were dissolved in 1M HCl for analysis with inductively coupled plasma atomic emission spectroscopy (Mo, Ni, V, Ba), Zeeman atomic absorption spectroscopy (Cd), and flow-injection analysis system (Se)¹³. Pyritic sulphur (S_{pyr}) was extracted by Cr(II) reduction¹⁶. Evolved sulphide was trapped in 1M NaOH and measured polarographically, or trapped as Ag_2S and measured for sulphur stable-isotope composition ($\delta^{34}S_{pyr}$) by combustion isotope ratio monitoring mass spectrometry relative to the V-CDT standard¹⁶. $R = {}^{34}S/{}^{32}S$, $\delta^{34}S_{pyr}(\text{‰}) = 1,000 \times [(R_{pyr}/R_{V-CDT}) - 1]$. Organic compounds were extracted with methanol/dichloromethane. Elemental sulphur was removed with activated Cu. Apolar fractions were isolated on activated Al_2O_3 , and analysed by gas chromatography and gas chromatography/mass spectrometry²⁰.

Received 17 July 1997; accepted 20 October 1998.

- Calvert, S. E., Nielsen, B. & Fontugne, M. R. Evidence from nitrogen isotope ratios for enhanced productivity during formation of eastern Mediterranean sapropels. *Nature* **359**, 223–225 (1992).
- Rosignol-Strick, M., Nesteroff, W., Olive, P. & Vergnaud-Grazzini, C. After the deluge: Mediterranean stagnation and sapropel formation. *Nature* **295**, 105–110 (1982).
- Rohling, E. J. Review and new aspects concerning the formation of eastern Mediterranean sapropels. *Mar. Geol.* **122**, 1–28 (1994).
- Emeis, K.-C. *et al.* Paleocceanography and sapropel introduction. In *Proc. ODP Init. Rep.* **Leg 160**, 21–28 (1996).
- Lourens, L. J. *et al.* Evaluation of the Plio-Pleistocene astronomical timescale. *Paleoceanography* **11**, 391–413 (1996).
- Liaaen-Jensen, S. in *The Photosynthetic Bacteria* (eds Clayton, R. K. & Sistrom, W. R.) 233–247 (Plenum, New York, 1978).
- Quandt, L., Gottschalk, G., Ziegler, H. & Stichter, W. Isotope discrimination by photosynthetic bacteria. *FEMS Microbiol. Lett.* **1**, 125–128 (1977).
- Summons, R. E. & Powell, T. G. Chlorobacteria in Palaeozoic seas revealed by biological markers, isotopes and geology. *Nature* **319**, 763–765 (1986).
- Kohnen, M. E. L. *et al.* Recognition of paleobiogeochemicals by a combined molecular sulfur and isotope geochemical approach. *Science* **256**, 358–362 (1992).
- Sinninghe Damsté, J. S., Wakeham, S. G., Kohnen, M. E. L., Hayes, J. M. & de Leeuw, J. W. A 6,000-year sedimentary molecular record of chemocline excursions in the Black Sea. *Nature* **362**, 827–829 (1993).
- Hartgers, W. A. *et al.* Evidence for only minor contributions from bacteria to sedimentary organic carbon. *Nature* **369**, 224–227 (1994).
- Repet, D. J., Simpson, D. J., Jørgensen, B. B. & Jannasch, H. W. Evidence for anoxygenic photosynthesis from the distribution of bacteriochlorophylls in the Black Sea. *Nature* **342**, 69–72 (1989).
- Nijenhuis, I. A., Brumsack, H.-J. & de Lange, G. J. The trace element budget of the Eastern Mediterranean during Pliocene sapropel formation. In *Proc. ODP Sci. Res.* **Leg 160** 199–206 (1998).
- Arthur, M. A., Brumsack, H.-J., Jenkyns, H. C. & Schlanger, S. O. In *Cretaceous Resources, Events and Rhythms* (eds Ginsburg, R. N. & Beaudoin, B.) 75–119 (Kluwer Academic, Dordrecht, 1990).
- Passier, H. F., Middelburg, J. J., van Os, B. J. H. & de Lange, G. J. Diagenetic pyritisation under eastern Mediterranean sapropels caused by downward sulphide diffusion. *Geochim. Cosmochim. Acta* **60**, 751–763 (1996).
- Passier, H. F., Middelburg, J. J., de Lange, G. J. & Böttcher, M. E. Modes of sapropel formation in the eastern Mediterranean: some constraints based on pyrite properties. *Mar. Geol.* **153**, 199–219 (1999).
- Passier, H. F., Middelburg, J. J., de Lange, G. J. & Böttcher, M. E. Pyrite contents, microtextures, and sulfur isotopes in relation to formation of the youngest eastern Mediterranean sapropel. *Geology* **25**, 519–522 (1997).
- Canfield, D. E., Lyons, T. W. & Raiswell, R. A model for iron deposition to euxinic Black Sea sediments. *Am. J. Sci.* **296**, 818–834 (1996).

- Kempe, S., Diercks, A.-R., Liebezeit, G. & Prange, A. in *Black Sea Oceanography* (eds Izdar, E. & Murray, J. W.) 89–110 (Kluwer Academic, Dordrecht, 1991).
- Canfield, D. E. & Thamdrup, B. The production of ^{34}S -depleted sulfide during bacterial disproportionation of elemental sulfur. *Science* **266**, 1973–1975 (1994).
- Fry, B. *et al.* Stable isotope studies of the carbon, nitrogen and sulfur cycles in the Black Sea and the Cariaco Trench. *Deep-Sea Res.* **38**, 1003–1019 (1991).
- Rosignol-Strick, M. Mediterranean Quaternary sapropels, an immediate response of the African monsoon to variation of insolation. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **49**, 237–263 (1985).
- Hilgen, F. J. Astronomical calibration of Gauss to Matuyama sapropels in the Mediterranean and implication for the geomagnetic polarity time scale. *Earth Planet. Sci. Lett.* **104**, 226–244 (1991).
- Arthur, M. A. *et al.* Varve calibrated records of carbonate and organic carbon accumulation over the last 2000 years in the Black Sea. *Glob. Biogeochem. Cycles* **8**, 195–217 (1994).
- Béthoux, J. P. Oxygen consumption, new production, vertical advection and environmental evolution in the Mediterranean Sea. *Deep-Sea Res.* **36**, 769–781 (1989).
- Dymond, J., Suess, E. & Lyle, M. Barium in deep-sea sediment: a geochemical proxy for paleo-productivity. *Paleoceanography* **7**, 163–181 (1992).
- Falkner, K. K. *et al.* The behavior of barium in anoxic marine waters. *Geochim. Cosmochim. Acta* **57**, 537–554 (1993).
- François, R., Honjo, S., Manganini, S. J. & Ravizza, G. E. Biogenic barium fluxes to the deep sea: Implications for paleoproductivity reconstruction. *Glob. Biogeochem. Cycles* **9**, 289–303 (1995).
- Dean, W. E., Gardner, J. V. & Piper, D. Z. Inorganic geochemical indicators of glacial-interglacial changes in productivity and anoxia on the California continental margin. *Geochim. Cosmochim. Acta* **61**, 4507–4518 (1997).
- Bosch, H.-J., Sinninghe Damsté, J. S. & de Leeuw, J. W. Molecular paleontology of Eastern Mediterranean sapropels: evidence for photic zone euxinia. In *Proc. ODP Sci. Res.* **Leg 160**, 285–295 (1998).

Acknowledgements. We thank S. McNab for linguistic advice; G. Nobbe, H. de Waard and T. Zalm for analytical assistance; and The Ocean Drilling Program and Leg 160 shipboard participants for their co-operation. This is a publication of the Netherlands Research School of Sedimentary Geology. This work was supported in part by the Netherlands Organization for Scientific Research (NWO) and the European Union Marine Science and Technology (MAST) programme.

Correspondence and requests for materials should be addressed to H.F.P. (e-mail: hpassier@geo.uu.nl).

Evidence for an early opening of the Bering Strait

Louie Marinovich Jr* & Andrey Yu. Gladenkov†

* Department of Invertebrate Zoology & Geology, California Academy of Sciences, Golden Gate Park, San Francisco, California 94118, USA

† Russian Academy of Sciences, Institute of the Lithosphere, Staromonetny per. 22, Moscow 109180, Russia

The first opening of the Bering Strait was an important palaeogeographical and biogeographical event for marine and terrestrial biotas in Asia and North America^{1–3}, and an oceanographic event of global importance^{4–6}. This event, however, has never been precisely dated, so it has not been accurately incorporated into models of global biogeography and oceanography. A recent find of the bivalve mollusc *Astarte* in southern Alaska is a clear signal that the strait had opened by at least the Late Miocene or earliest Pliocene epochs, because this genus otherwise dwelled only in the Arctic and North Atlantic oceans at that time^{1,2}. Here we show that marine diatoms with *Astarte* are correlative with subzone b of the *Neodenticula kamtschatica* zone of the North Pacific diatom biochronology, yielding an age range of 4.8–5.5 Myr (refs 7, 8). Stratigraphic information suggests that this may be a minimum age range for the strait's first opening, which evidently occurred between 4.8 and 7.3–7.4 Myr ago. Our results contrast with past studies^{9–13} that suggested an age of 3.1–4.1 Myr for the initial opening of the Bering Strait.

It would be difficult to name a region as significant for palaeogeography or biogeography as the Bering Strait. A closed strait joined Eurasia and North America into one supercontinent and allowed the interchange of terrestrial animals and plants, but sealed off the marine connection between the North Pacific and the North Atlantic-Arctic oceans¹⁴. Mollusc faunas in these two oceanic realms evolved separately, and taxa in the Arctic Ocean Cenozoic marine biota were either endemic or of North Atlantic affinities before the Bering Strait formed^{13–17}. The opening of the strait produced a dramatic change in the composition of Arctic, North Atlantic and North Pacific shallow-water marine faunas^{1,2,11,18,19}. Much of the modern Arctic Ocean molluscan fauna, particularly in the North American Arctic and along the eastern coast of Siberia, is dominated by species of Pacific origin^{2,18}. The first opening of the Bering Strait

temporarily ended terrestrial migrations between Asia and North America^{3,20}, and allowed marine organisms to migrate between the Arctic and North Pacific oceans for the first time since the middle Cretaceous period (Albian stage, about 105 Myr ago)¹⁴. Despite the importance of the Bering Strait's first opening, determining its age has proved difficult, mostly due to the apparent absence of a stratigraphic sequence with evidence of an open strait and age-diagnostic fossils.

The age of the Bering Strait was conjectural before it was tentatively assigned to the Late Miocene, based on a comparison of Arctic Ocean and North Pacific Neogene molluscan faunas¹⁵. However, the Late Miocene subepoch extended from 11.2 to 5.32 Myr ago²¹, so the possible age of the strait varied by a factor of more than two. More recent workers have placed the opening of the Bering Strait in the middle Pliocene at 3.1 Myr (refs 9, 10), 3.6 Myr (refs 19, 22), or 4.1 Myr (refs 12, 13), when adjusted to the timescale²¹ used here, based on occurrences of Pacific molluscs in the Arctic Ocean⁹ or North Atlantic^{1,23}. The appearance of North Pacific species in the North Atlantic is shown especially well in the Pliocene and Pleistocene Tjörnes sequence of Iceland, where molluscs such as *Clinocardium ciliatum*, *Mya*, *Musculus*, *Serripes groenlandicus*, *Amphissa*, *Littorina*, *Natica clausa*, *Neptunea* and *Nucella*, which have well-documented Miocene or older histories in the North Pacific¹, make up some 25% of the so-called *Serripes* zone molluscan fauna²³. Pillow lavas within the *Serripes* zone occur stratigraphically just above the first appearance of North Pacific molluscs, and are dated at "more than 3.35" Myr on the basis of palaeomagnetic studies^{23,24}. The time of opening of the Bering Strait was thought to be slightly older than these pillow basalts²⁴.

The earliest known evidence in the Pacific for an open Bering Strait is the presence of the Atlantic-Arctic bivalve *Astarte* in Neogene faunas of southern Alaska²⁵, Kamchatka^{26,27} and northern Japan^{28,29}. The best-documented age of 4.05 Myr was based on *Astarte* in the upper part of the Limimtevayam Formation on Karaginsky Island, eastern Kamchatka^{26,27} (Fig. 1), in beds assigned to the lowermost part of the *Neodenticula koizumii*-*Neodenticula kamtschatica* diatom zone (with its base at 3.4–4.1 Myr; ref. 7) (Fig. 2), and was modified to 4.4 Myr (ref. 8) when adjusted to the newer timescale²¹ used here.

A stratigraphically lower occurrence of *Astarte* in the lowermost part of the Limimtevayam Formation²⁶, just above its unconform-

able contact with the underlying Yununvayam Formation, is not well dated. Diatoms are absent from both the upper part of the Yununvayam Formation and from the overlying *Astarte*-bearing basal Limimtevayam Formation²⁷. The lower part of the Yununvayam Formation contains a diatom assemblage correlative with subzone a of the *N. kamtschatica* zone (5.5 Myr to 7.3–7.4 Myr; refs 7, 8) (Fig. 2). The stratigraphically much higher beds in the upper part of the Limimtevayam Formation contain diatoms correlative with the upper part of the *N. kamtschatica* zone, most likely with its subzone c (3.4–4.1 to 4.8 Myr). Therefore, the maximum possible age for the stratigraphically lowest *Astarte* known on Karaginsky Island is 7.3–7.4 Myr, the base of subzone a of the *N. kamtschatica* zone. In summary, diatoms in the Karaginsky Island sequence are absent from a biostratigraphic interval beginning somewhere in subzone a of the *N. kamtschatica* zone (with a maximum age of 7.3–7.4 Myr), to somewhere in subzone c of the *N. kamtschatica* zone (3.4–4.1 to 4.8 Myr). This evidence implies a maximum possible age range of 4.8 to 7.3–7.4 Myr for the oldest Pacific Cenozoic *Astarte* and for the initial opening of the Bering Strait.

A sequence of molluscs and diatoms similar to that on Karaginsky Island occurs in the Bear Lake Formation of the Alaska Peninsula, southwestern Alaska (Fig. 1). *Astarte borealis* and *Astarte* sp. occur with diatoms at two horizons in the 276-m-thick Sandy Ridge stratigraphic section³⁰ (Figs 1, 2). A flora of more than 100 marine diatom taxa contains age-diagnostic species (Table 1) that represent subzone b of the North Pacific *N. kamtschatica* zone (Fig. 2), which has an age range of 4.8–5.5 Myr. The presence of *Thalassiosira oestrupii* and *T. insigna* at Sandy Ridge is particularly important, because the base of subzone b is defined by the first appearance of *T. oestrupii* at 5.5 Myr, and its top is defined by the last appearance of *T. insigna* at 4.8 Myr (refs 7, 31). This may be a minimum age range for the strait's first opening, because the diatom samples are thought to come from the upper part of the stratigraphic section. Geological fieldwork at Sandy Ridge indicates that much of the *Astarte*-bearing section lies downsection and may be significantly older. The Miocene/Pliocene boundary is currently²¹ placed at 5.32 Myr, so molluscan and diatom evidence from Alaska and Kamchatka suggests that the Bering Strait first formed in the Late Miocene or earliest Pliocene, between 4.8 and 7.3–7.4 Myr.

The variety of ages previously cited for the Bering Strait shows that even Arctic geologists and palaeontologists have not had a

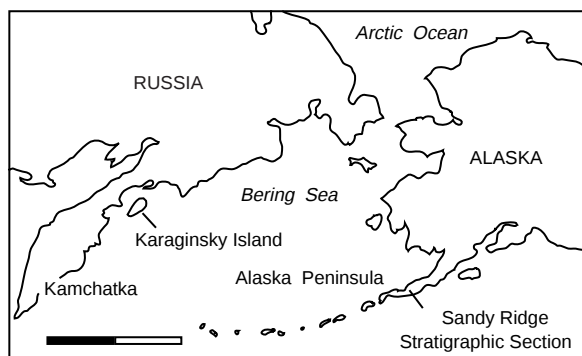


Figure 1 Map showing locations of places mentioned in the text. Karaginsky Island, northeastern Kamchatka, and the Sandy Ridge section on the Alaska Peninsula, southwestern Alaska, are the two sites where molluscan evidence of an open Bering Strait, the bivalve mollusc *Astarte*, occurs with age-diagnostic marine diatoms. Scale bar, 1,000 km.

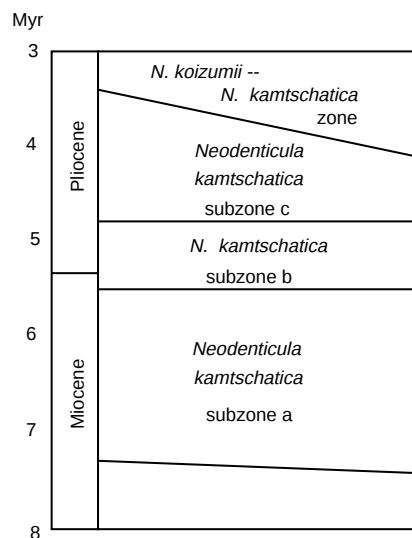


Figure 2 Diatom zonation for the high-latitude North Pacific³, revised with an updated timescale²¹. Geological ages in millions of years are on the left margin. Marine diatoms suggest that the first opening of the Bering Strait took place sometime just after the beginning of *Neodenticula kamtschatica* subzone a at 7.3–7.4 Myr and before the end of *N. kamtschatica* subzone b at 4.8 Myr.

Table 1 Diatoms from the Bear Lake Formation at Sandy Ridge

<i>Actinocyclus ochotensis</i>	<i>Thalassiosira antiqua</i>
<i>Bacterosira fragilis</i>	<i>T. decipiens</i>
<i>Cymatosira debyi</i>	<i>T. dolmatovae</i>
<i>Delphineis sachalinensis</i>	<i>T. eccentrica</i>
<i>D. simonsenii</i>	<i>T. gravida</i>
<i>Detonula confervacea</i>	<i>T. hyalina</i>
(syn.: <i>Melosira albicans</i>)	<i>T. insigna</i>
<i>Lithodesmium minusculum</i>	<i>T. jouseae</i>
<i>Neodenticula kamtschatica</i>	<i>T. oestrupii</i>
<i>Nitzschia cylindra</i>	<i>T. orientalis</i>
<i>N. grunowii</i>	<i>T. praeoestrupii</i>
<i>Porosira glacialis</i>	<i>T. sheshukovae</i>
<i>Pyxidicula zabelinae</i>	(syn.: <i>Pseudopodosira elegans</i>)
<i>Rhaphoneis angularis</i>	<i>Thalassiothrix robusta</i>
<i>Thalassionema nitzschioides</i>	<i>Trochosira concava</i>

consistent conceptual framework that accurately incorporates this age into global geological and palaeontological history. An accurate date for this event will allow it to be incorporated into the history of terrestrial mammal migrations between Eurasia and North America^{3,10}, and into the history of marine migrations between the North Pacific and the Atlantic-Arctic oceans^{2,11,19}. The arrival of Atlantic-Arctic *Astarte* in the Pacific before 4.8 Myr contrasts with the much later appearance of Pacific molluscs in the North Atlantic at about 3.5 Myr (refs 2, 24). This disparity may reflect a change in the dominant direction of water flow and faunal migration through the Bering Strait. A reversed (that is, southward) flow through the Bering Strait is one consequence of an open Miocene Isthmus of Panama in experiments with ocean general circulation models⁴. A dominant southward flow through the Bering Strait may have taken place from its first opening at 4.8–5.5 Myr (or earlier) and continued until sometime after 4.6 Myr, when a critical threshold in the closure history of the Isthmus of Panama caused a marked reorganization of Northern Hemisphere ocean circulation³², including the onset of the present northward flow through the Bering Strait^{4,32}. The well-documented invasion of the Arctic and North Atlantic oceans by Pacific molluscs^{1,2,11,14,23} evidently resulted from a change to northward flow through the Bering Strait. Continuing research on molluscan events and microfossil chronostratigraphy in Alaska is expected to refine the age of the first opening of the Bering Strait and resulting biotic events, and to relate them more precisely to global oceanographic history. □

Received 18 September; accepted 19 October 1998.

- Durham, J. W. & MacNeil, F. S. in *The Bering Land Bridge* (ed. Hopkins, D. M.) 326–349 (Stanford Univ. Press, 1967).
- Vermeij, G. J. Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology* **17**, 281–307 (1991).
- Woodburne, M. O. & Swisher, C. C. in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V., Aubry, M. P. & Hardenbol, J.) 336–364 (Spec. Publ. 54, SEPM (Soc. for Sedimentary Geology), Tulsa, 1995).
- Maier-Reimer, E. & Mikolajewicz, U. Ocean general circulation model sensitivity experiment with an open Central American isthmus. *Paleoceanography* **5**, 349–366 (1990).
- Shaffer, G. & Bendtsen, J. Role of the Bering Strait in controlling North Atlantic ocean circulation and climate. *Nature* **367**, 354–357 (1994).
- Reason, C. J. C. & Power, S. B. The influence of the Bering Strait on the circulation in a coarse resolution global ocean model. *Clim. Dyn.* **9**, 363–369 (1994).
- Barron, J. A. & Gladenkov, A. Y. *Proc. ODP Leg 145* 3–19 (1995).
- Marincovich, L. Jr & Gladenkov, A. Y. in *Beringian Paleoenvironments Workshop, Program and Abstracts* (eds Brigham-Grette, J. & Elias, S.) 103–104 (Florissant, Colorado, 1997).
- Fyles, J. G., Marincovich, L. Jr, Matthews, J. V. Jr & Barendregt, R. Unique mollusc find in the Beaufort Formation (Pliocene) on Meighen Island Arctic Canada. *Geol. Surv. Can. Curr. Res. Pap.* **91–1B**, 105–112 (1991).
- Repenning, C. A. & Brouwers, E. M. Late Pliocene-early Pleistocene ecological changes in the Arctic Ocean Borderland. *US Geol. Surv. Bull.* **2036**, (1992).
- Vermeij, G. J. Invasion and extinction: the last three million years of North Sea pelecypod history. *Conserv. Biol.* **3**, 274–281 (1989).
- Brigham-Grette, J., Carter, L. D., Marincovich, L., Brouwers, E. & Hopkins, D. M. *Pliocene High-latitude Climate Records* (ed. Ishman, S. E.) 5–6 (Open-file Rep. 94-588) (US Geol. Surv., Menlo Park, California, 1994).
- Nolf, D. & Marincovich, L. Jr First record of fossil *Merlangius* (Pisces, Gadiformes) from arctic Alaska and chronostratigraphic implications. *Contrib. Tert. Quatern. Geol.* **31**, 9–13 (1994).
- Marincovich, L. Jr, Brouwers, E. M., Hopkins, D. M. & McKenna, M. C. in *The Geology of North America Vol. L, The Arctic Ocean Region* 403–426 (Geol. Soc. Am., Boulder, Colorado, 1990).
- MacNeil, F. S. Cenozoic megafossils of northern Alaska. *US Geol. Surv. Prof. Pap.* **294-C**, 99–126 (1957).
- Marincovich, L. Jr Danian mollusks from the Prince Creek Formation, northern Alaska, and implications for Arctic Ocean paleogeography. *Paleontol. Soc. Mem.* **35**, (1993).
- Marincovich, L. Jr & Zinsmeister, W. J. The first Tertiary (Paleocene) marine mollusks from the Eureka Sound Group, Ellesmere Island, Canada. *J. Paleontol.* **65**, 242–248 (1991).

- Evseev, G. A. & Krasnov, E. V. in *Beringia in the Cenozoic Era* (ed. Kontrimavichus, V. L.) 50–64 (Acad. Sci. USSR, Far Eastern Sci. Center, Vladivostok, 1984).
- Vermeij, G. J. Geographical restriction as a guide to the causes of extinction: the case of the cold northern oceans during the Neogene. *Paleobiology* **15**, 335–356 (1989).
- Repenning, C. A. Faunal exchanges between Siberia and North America. *Can. J. Anthropol.* **1**, 37–44 (1980).
- Berggren, W. A., Kent, D. V., Swisher, C. C. & Aubry, M.-P. in *Geochronology, Time Scales and Global Stratigraphic Correlations* (eds Berggren, W. A., Kent, D. V., Aubry, M.-P. & Hardenbol, J.) 129–212 (Spec. Publ. 54, SEPM (Soc. for Sedimentary Geology), Tulsa, 1995).
- Herman, Y. & Hopkins, D. M. Arctic oceanic climate in Late Cenozoic time. *Science* **209**, 557–562 (1980).
- Gladenkov, Y. B., Norton, P. & Spaink, G. Upper Cenozoic of Iceland (stratigraphy of Pliocene-Pleistocene and paleontological assemblages). *U.S.S.R. Acad. Sci. Trans.* **345**, (1980). (In Russian.)
- Einarsson, T., Hopkins, D. M. & Doell, R. D. in *The Bering Land Bridge* (ed. Hopkins, D. M.) 312–325 (Stanford Univ. Press, 1967).
- Addicott, W. O., Winkler, G. R. & Plafker, G. *Preliminary Megafossil Biostratigraphy and Correlation in the Gulf of Alaska Tertiary Province* (Open-file Rep. 78-491) (US Geol. Surv., Menlo Park, California, 1978).
- Gladenkov, Y. B. The Neogene of Kamchatka (problems of Biostratigraphy and Paleogeology). *USSR Acad. Sci. Trans.* **214**, (1972). (In Russian.)
- Gladenkov, Y. B., Barinov, K. B., Basilian, A. E. & Cronin, T. M. Stratigraphy and paleoceanography of Miocene deposits of Karaginsky Island, eastern Kamchatka. *USSR Quart. Sci. Rev.* **10**, 239–245 (1991).
- Uozumi, S., Akamatsu, M. & Takagi, T. in *Japanese Cenozoic Molluscs—Their Origin and Migration* (eds Kotaka, T. & Marincovich, L. Jr) 211–226 (Spec. Pap. 29, Palaeontol. Soc. Japan, Tokyo, 1986).
- Suzuki, A. & Akamatsu, M. Post-Miocene cold-water molluscan faunas from Hokkaido, Northern Japan. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **108**, 353–367 (1994).
- Detterman, R. L., Case, J. E., Miller, J. W., Wilson, F. H. & Yount, M. E. Stratigraphic framework of the Alaska Peninsula. *US Geol. Surv. Prof. Pap.* **1969-A**, (1996).
- Barron, J. A. in *The Centenary of Japanese Micropaleontology* (eds Ishizaki, K. & Saito, T.) 413–425 (Tokyo Univ. Press, 1992).
- Haug, G. H. & Tiedemann, R. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature* **393**, 673–676 (1998).

Acknowledgements. A. Yu. Gladenkov was supported by the Russian Foundation for Fundamental Research. We thank the California Academy of Sciences for providing mollusc samples for diatom analysis.

Correspondence and requests for materials should be addressed to L.M. (e-mail: lmarin@best.com).

Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee

Boris Baer & Paul Schmid-Hempel

ETH Zürich, Experimental Ecology, ETH-Zentrum NW, CH-8092 Zürich, Switzerland

In many species of animals, females typically mate with more than one male (polyandry). Some social insects carry this behaviour to extremes¹. For example, honeybee queens mate with ten to twenty (or even more) males on their nuptial flights². The reasons for this behaviour remain unknown, given the obvious costs of time, energy and exposure to predation. Several potential benefits of polyandry have been proposed^{1,3,4}, but none are well supported yet. Here we test the hypothesis that genetic diversity among a female's offspring may offer some protection from parasitism^{5–7}. We artificially inseminated queens of a bumble-bee (*Bombus terrestris* L.) with sperm of either low or high genetic diversity. The resulting colonies were exposed to parasitism under field conditions. High-diversity colonies had fewer parasites and showed greater reproductive success, on average, than did low-diversity colonies. We suggest that female mating frequency may be influenced in part by parasites.

The diversity versus parasites hypothesis assumes that there are genotypic interactions between hosts and parasites^{8–10} and predicts that, in a parasite-challenged population, genetically diverse colonies should have a higher probability of survival and/or higher reproductive success^{7,11}. Increased genetic diversity reduces parasite transmission among bumble-bee workers⁹. Therefore, the fraction of workers infected by a given parasite should be lower in genetically diverse colonies.

Polyandry occurs in a number of bumble-bee species¹². Bumblebees (*Bombus* spp.) produce annual colonies founded by single

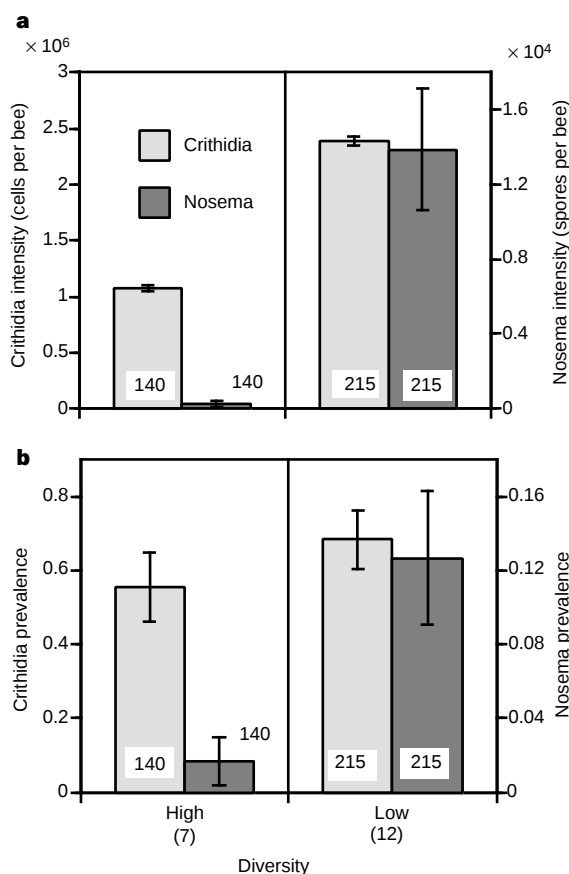


Figure 1 Intensity and prevalence of infection of high- and low-diversity *B. terrestris* colonies. **a**, Average infection intensities were lower in workers from high-diversity colonies (left) than in workers from low-diversity colonies (right) for the two major infectious parasites, *C. bombi* (Mann-Whitney *U*-test: $z = 2.62$, $P = 0.005$) and *N. bombi* ($z = 2.503$, $P = 0.006$). Infection levels are in a range typically observed in field samples and known to produce serious effects. **b**, Prevalence of infection (the proportion of parasitized workers in a colony) was significantly lower in high-diversity colonies than in low-diversity colonies for *N. bombi* (*U*-test: $z = 2.327$, $P = 0.01$). Prevalence was also lower for *C. bombi* but not significantly so ($z = 1.144$, $P = 0.12$). Numbers within bars indicate the number of workers checked per group. Number of colonies per treatment group are given in parentheses at the bottom. Values are means $\pm$ s.e.m.

queens in spring. As the colony's work-force grows, the workers forage and become exposed to parasites¹³. At the end of the colony cycle reproduction occurs, with the production of sexuals (drones and young queens) that mate, after which the fertilized queens overwinter and the males die. *B. terrestris* provides a good model system for the study of the possible consequences of variation in mating frequency and parasitism. It shows biology typical of the genus and can be bred easily in the laboratory, enabling the experimental manipulation of queen mating frequency. Although queens of *B. terrestris* may be mostly singly mated in natural populations of central Europe¹⁴, multiple mating has been reported^{12,15}. Furthermore, *B. terrestris* may exhibit geographic variation in mating frequency, as indicated for *B. hypnorum* in central¹⁴ compared with northern Europe¹⁶. Finally, *B. terrestris* has been extensively studied as a host for a wide range of parasites, including protozoa, mites, wax moths and parasitoid flies^{17,18}. Infection by such parasites reduces colony performance and reproductive success^{19–21}.

We manipulated the amount of polyandry by artificially inseminating queens with sperm from drones to generate colonies of low genetic diversity (corresponding to an effective mating frequency of $n_e \approx 1.33$, that is, very close to single mating; twelve colonies) or

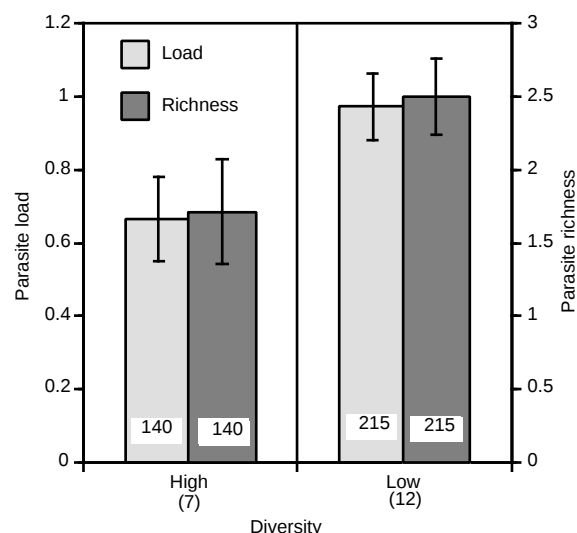


Figure 2 Average parasite load (the average number of different parasite species per worker per colony) was lower in workers from high-diversity colonies than in workers from low-diversity colonies (*U*-test: $z = 2.162$, $P = 0.015$). Similarly, parasite richness (the number of different parasite species per colony) was lower in high-diversity colonies ($z = 1.694$, $P = 0.045$). Sample sizes are as explained in Fig. 1. Values are means $\pm$ s.e.m.

high genetic diversity ($n_e \approx 4$, as for multiple mating; seven colonies) (see Methods). Because the effective mating frequency was independent of any possible variation in male quality (such as susceptibility to parasites), the experimental effect was merely to increase genetic diversity within the colony rather than to select for 'good genes'.

From worker samples we detected the following parasites in our colonies: *Crithidia bombi*, *Nosema bombi*, *Mattesia (Apicystis) bombi* (gregarina), mites (Acari) and endoparasitic larvae of conopid flies (Conopidae, Diptera). Colonies headed by queens inseminated by highly diverse sperm had lower intensities of infection and lower prevalence of infection, as compared with colonies with low genetic diversity, by both of the major parasites seen in our area of Switzerland (Fig. 1). Parasite load and parasite richness were also lower in genetically diverse colonies (Fig. 2). High-diversity colonies attained somewhat larger sizes (mean $\pm$ s.e.m., 52.0 ± 15.8 workers; $n = 7$ colonies) than low-diversity colonies (32.2 ± 4.6 ; $n = 12$) but, because of typically high variation in colony size, this difference was not significant ($t = 1.49$, degree of freedom (d.f.) = 17, $P = 0.08$). Finally, the high-diversity colonies had higher reproductive success, as assessed by several criteria (Fig. 3) including the total number of sexual organisms produced ($t = 2.57$, d.f. = 17, $P = 0.01$). Genetically diverse colonies therefore performed better and were subject to less parasitism.

Our results are direct evidence that polyandry by females lowers parasitism under natural conditions, as implied by earlier findings from the same and other species^{9,22}. Although we cannot infer the precise mechanisms by which parasitism is reduced, the effect seems to involve increased genetic diversity among workers within the colony^{23,24}. Hence, future analyses of mating strategies in animals should consider the benefits of genetic variability among offspring in relation to exposure to parasites, especially in philopatric species in which offspring will be exposed to similar conditions to their parents, or in which parents and offspring cooperate in breeding and may therefore acquire the same diseases.

At first sight, our results also present a paradox in that it is not known whether multiple mating by queens of *B. terrestris* is common in any population¹⁴. This indicates that queens do not or cannot take advantage of the substantial benefits of mating with

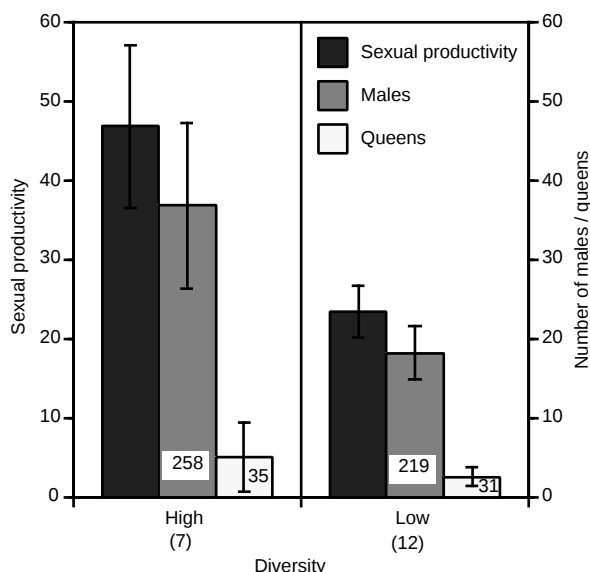


Figure 3 Average reproductive success was higher in high-diversity colonies than in low-diversity colonies, as assessed by several criteria: first, the males produced ($t = 2.05$, d.f. = 17, $P = 0.028$); second, the number of queens produced (U -test, $z = 0.518$, not significant); and third, sexual productivity (that is, the number of queens multiplied by two, plus the number of males; justified because the body mass of queens is approximately double that of males: $t = 2.67$, d.f. = 17, $P = 0.008$). Sample sizes are explained in Fig. 1 with numbers inside bars representing the total number of sexual organisms sampled. Values are means $\pm$ s.e.m.

several males. The time, energy or predation risk involved in soliciting additional matings alone are unlikely to prevent polyandry, although the relevant data do not exist. Instead, we suggest that male–male competition may have led to strategies that render a second mating impossible, or at least unprofitable, for the female. We recently found that male *B. terrestris* deposit large amounts of a sticky secretion (a mating plug, previously unknown for the Bombini) in the female's genital duct after they have transferred their sperm, thus possibly preventing further insemination by other males for a period of at least two days (N. Duvoisin, B.B. and P.S.-H., manuscript in preparation). Hence, a conflict of interest may deny the benefits of polyandry for the females of *B. terrestris*. □

Methods

Experimental protocol. To inseminate queens, we dissected males and collected sperm from accessory testes. Queens were briefly anaesthetized and the sperm were injected into the reproductive tract. This procedure led to the production of normal colonies. Subsequent microsatellite analysis of worker genotypes in the experimental colonies confirmed that the inseminations were effective and that the genotypes of the drones were present in roughly equal proportions. A detailed description of the insemination technique will be published elsewhere. Because of current technical limitations in the procedure, we used the combined sperm from four males to inseminate a given queen. For the high-diversity group (seven colonies), we used sperm from four drones, each from one of four unrelated colonies. For the low-diversity group (twelve colonies)—intended to mimic low paternity frequency—we used sperm from four brothers of a given colony (males are haploid and brothers are related to each other by one-half), therefore also controlling for the number of contributors and the amount of sperm. In addition, drones and queens were assigned to one another such that mother colonies of drones were represented with equal chances both in the low- and the high-diversity group. In all cases, we inseminated with sperm from drones that were not related to the queen. Queens were otherwise randomly allocated to the two treatment groups, high or low diversity. This procedure resulted in an estimated among-worker relatedness (and effective mating frequency, n_e) of $g = 0.375$ ($n_e \sim 4$) and

$g = 0.625$ ($n_e \sim 1.33$)¹ for the high- and low-diversity groups, respectively. The inseminated queens were overwintered at 6 °C for one month before they were allowed to start a colony in the laboratory. As soon as the second brood had hatched we transferred the colony to our field site, a typical bumble-bee habitat in an area of flowering meadows near Basel, Switzerland. Transfers were matches so that date of field placement and microlocality were balanced between the two treatment groups. Although the early part of the colony cycle was not in the field, the colonies nevertheless developed for much of their cycle under natural conditions. We have no indication that this experimental limitation affects the conclusions. The phenology of the experimental colonies was the same as that of the wild colonies in the area.

Data collection and analysis. After their placement in the field, the colonies were left undisturbed except for regular checks during the night hours every 7 days. On these occasions, we checked the colonies for general status, counted the number of workers present and took a random sample of ~10% of the workers, which were immediately freeze-killed for later study in the laboratory. This fraction is small compared with naturally occurring losses of workers. Towards the end of the colony cycle, we checked the colonies more often and also collected the newly emerged sexual individuals to assess reproductive success. Observers who studied the workers for parasites were blind to the origin of the samples. We checked the bees' external surfaces, gut, Malpighian tubules, fat body and other parts of the haemocoel for parasites. Concentrations of cells of *C. bombi* and spores of *N. bombi* were measured in a haemocytometer (Neubauer chamber). Measures of parasitism were as follows: first, parasite richness, that is, the number of different parasite species per colony; second, parasite load, that is, the average number of different parasite species per worker per colony; third, prevalence of infection, that is, the proportion of parasitized workers in a colony; and fourth, intensity of infection, that is, the estimated number of parasites within a given host individual. During the year of our study, only *C. bombi* and *N. bombi* were common enough to allow us to calculate meaningful values for prevalence and intensity. Parasite richness and load were calculated for all parasites found.

We analysed a total of 358 workers from both treatment groups. All experimental colonies survived to the end of the season. We found no or only a weakly positive relationship between colony size or the absolute number of workers checked and our estimates of parasitism in the colony. Measures of parasitism were not corrected for effects of sample size. As more workers were sampled and checked from high-diversity colonies (which were on average somewhat larger), any potential sampling bias, and also any unrelated effect of colony size, would have run counter to our hypothesis. Our analysis of the treatment effect is, therefore, conservative. Data were analysed with program SPSS6.1. All tests are reported with one-tailed probabilities, as the hypothesis predicts a directional difference in favour of high-diversity colonies.

Received 6 May; accepted 20 October 1998.

- Boomsma, J. J. & Ratnieks, F. L. W. Paternity in eusocial hymenoptera. *Phil. Trans. R. Soc. Lond. B* **351**, 947–975 (1996).
- Moritz, F. R. A. et al. High degree of polyandry in *Apis dorsata* queens detected by DNA microsatellite variability. *Behav. Ecol. Sociobiol.* **37**, 357–363 (1995).
- Boomsma, J. J. & Grafen, A. Colony-level sex ratio selection in the eusocial Hymenoptera. *J. Evol. Biol.* **3**, 383–407 (1991).
- Keller, L. & Reeve, H. K. Partitioning of reproduction in animal societies. *TREE* **9**, 98–102 (1994).
- Tooby, J. Pathogens, polymorphism, and the evolution of sex. *J. Theor. Biol.* **97**, 557–576 (1982).
- Hamilton, W. D. in *Animal Societies: Theory and Facts* (eds Itô, Y., Brown, J. L. & Kikkawa, J.) 81–102 (Japan. Sci. Soc. Press, Tokyo, 1987).
- Sherman, P. W., Seeley, T. D. & Reeve, H. K. Parasites, pathogens, and polyandry in social Hymenoptera. *Am. Nat.* **131**, 602–610 (1988).
- Kulincevic, J. M. in *Bee Genetics and Breeding* (ed. Rinderer, T. E.) 391–414 (Academic, Orlando, 1986).
- Shykoff, J. A. & Schmid-Hempel, P. Parasites and the advantage of genetic variability within social insect colonies. *Proc. R. Soc. Lond. B* **243**, 55–58 (1991).
- Schmid-Hempel, P. & Loosli, R. A contribution to the knowledge of *Nosema* infections in bumble bees, *Bombus* spp. *Apidologie* (in the press).
- Schmid-Hempel, P. Infection and colony variability in social insects. *Phil. Trans. R. Soc. Lond. B* **346**, 313–321 (1994).
- Crozier, R. H. & Pamilo, P. *Evolution of Social Insect Colonies* 1–306 (Oxford Univ. Press, 1996).
- Durrer, S. & Schmid-Hempel, P. Shared use of flowers leads to horizontal pathogen transmission. *Proc. R. Soc. Lond. B* **258**, 299–302 (1994).
- Estoup, A., Scholl, A., Pouvreaux, A. & Solignac, M. Monandry and polyandry in bumble bees (Hymenoptera, Bombinae) as evidenced by highly variable microsatellites. *Mol. Ecol.* **4**, 89–93 (1995).
- Röseler, P.-F. Die Anzahl Spermien im Receptaculum seminis von Hummelköniginnen (Hymenoptera, Apidae, Bombinae). *Apidologie* **4**, 267–274 (1973).
- Thorén, P., Estoup, A., Paxton, R. J. & Tengö, J. in *6th Congr. Eur. Soc. Evol. Biol. Arnhem* S3 (Eur. Soc. Evol. Biol., Arnhem, 1997).
- MacFarlane, R. P., Lipa, J. J. & Liu, H. J. Bumble bee pathogens and internal enemies. *Bee World* **76**, 130–148 (1995).

18. Schmid-Hempel, P. *Parasites in Social Insects* (Princeton Univ. Press, Princeton, 1998).
19. Shykoff, J. A. & Schmid-Hempel, P. Parasites delay worker reproduction in bumblebees: consequences for eusociality. *Behav. Ecol.* **2**, 242–248 (1991).
20. Müller, C. B. & Schmid-Hempel, P. Variation in worker mortality and reproductive performance in the bumble bee, *B. lucorum*. *Funct. Ecol.* **6**, 48–56 (1992).
21. Müller, C. B. & Schmid-Hempel, P. Correlates of reproductive success among field colonies of *Bombus lucorum* L.: the importance of growth and parasites. *Ecol. Entomol.* **17**, 343–353 (1993).
22. Liersch, S. & Schmid-Hempel, P. Genetic variability within social insect colonies reduces parasite load. *Proc. R. Soc. Lond. B* **265**, 221–225 (1998).
23. Walker, W. F. Sperm utilization strategies in non-social insects. *Am. Nat.* **115**, 780–799 (1980).
24. Madsen, T., Shine, R., Loman, J. & Håkansson, T. Why do female adders copulate so frequently? *Nature* **355**, 440–441 (1992).

Acknowledgements. We thank K. Boomsma, M. Brown, F. Fischer, B. Imhoof, S. Koulanos, N. Krüger, E. Magro, E. Meier, R. E. Page Jr, F. Ratnieks, R. Schmid-Hempel and H. Schwarz for help and comments, and the local communities and forestry commissions for use of their land. This work was supported by grants to P.S.-H. from the Swiss National Science Foundation and the Swiss Office of Science and Technology (within a European TMR network).

Correspondence and requests for materials should be addressed to P.S.-H. (e-mail: psh@eco.unmnw.ethz.ch).

Responses of auditory-cortex neurons to structural features of natural sounds

Israel Nelken, Yaron Rotman & Omer Bar Yosef

Department of Physiology, Hadassah Medical School, Ein Karem, PO Box 12272, Jerusalem 91120, Israel

Sound-processing strategies that use the highly non-random structure of natural sounds may confer evolutionary advantage to many species. Auditory processing of natural sounds has been studied almost exclusively in the context of species-specific vocalizations^{1–4}, although these form only a small part of the acoustic biotope⁵. To study the relationships between properties of natural soundscapes and neuronal processing mechanisms in the auditory system, we analysed sound from a range of different environments. Here we show that for many non-animal sounds

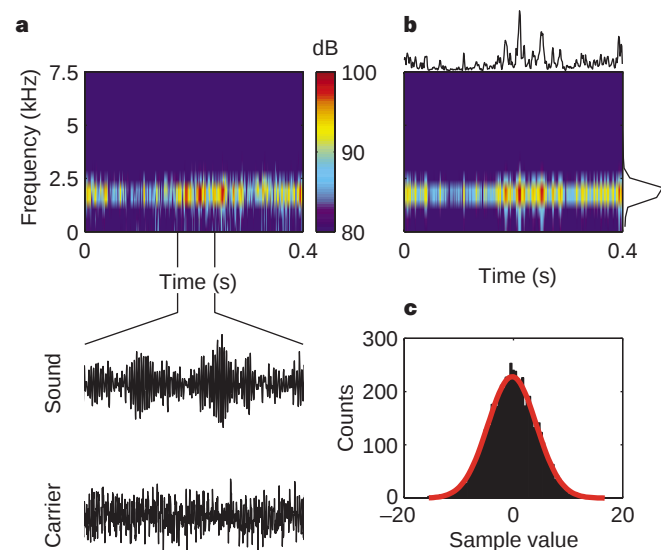


Figure 1 An example of the separation of a sound into a separable spectrogram and a carrier. **a**, The spectrogram together with the temporal waveform of a segment of the sound and the demodulated carrier. **b**, The separable approximation. The temporal and spectral components are shown at the margins. Same scale as in **a**. **c**, The amplitude distribution of the carrier, with a fit to the gaussian distribution ($\chi^2 = 74$, d.f. = 75, not significant). The amplitude distribution of the original sound segment is clearly non-gaussian ($\chi^2 = 222$, d.f. = 75, $P = 0$).

and background mixtures of animal sounds, energy in different frequency bands is coherently modulated. Co-modulation of different frequency bands in background noise facilitates the detection of tones in noise by humans, a phenomenon known as co-modulation masking release (CMR)^{6,7}. We show that co-modulation also improves the ability of auditory-cortex neurons to detect tones in noise, and we propose that this property of auditory neurons may underlie behavioural CMR. This correspondence may represent an adaptation of the auditory system for the use of an attribute of natural sounds to facilitate real-world processing tasks.

We analysed soundscapes (long recordings of mixtures of animal vocalizations and non-animal sounds) and vocalizations of single animals (for example, frogs, birds, cats and dogs). More than 1300 s from 25 sources were analysed. Following previous work on natural images⁸, the sounds were decomposed into an envelope modulating a gaussian carrier. The spectrogram was used as a spectrotemporal envelope. The parameters of the spectral analysis were fixed at values that resulted in the carrier having gaussian statistics for most sound segments.

Figure 1 shows the spectrogram of a segment from a chorus of crows, which is clearly dominated by a fluctuating horizontal stripe. This structure indicates that the spectrogram is approximately separable: it is well approximated by a product of a function of

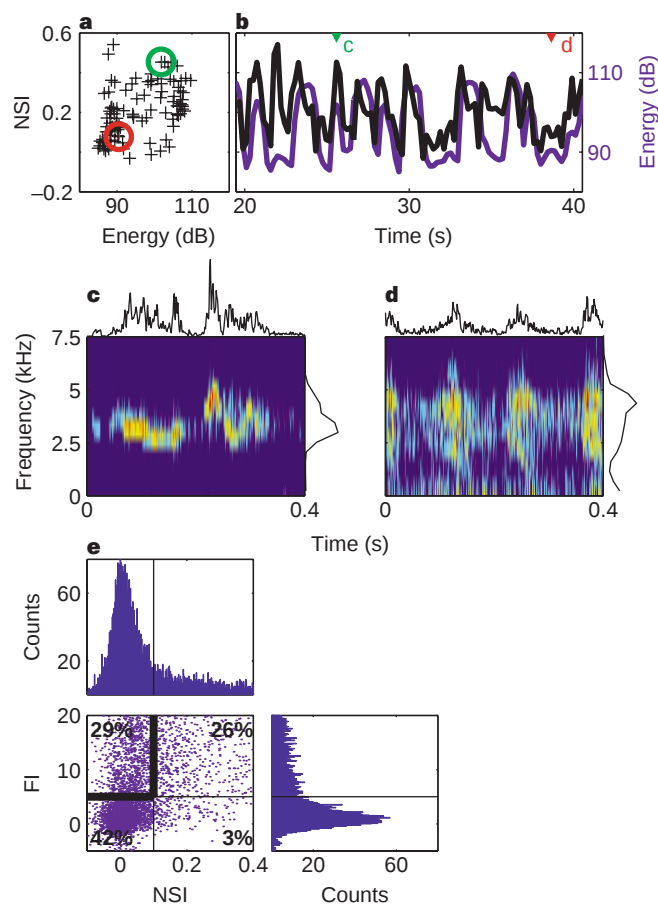


Figure 2 Statistics of soundscapes. **a**, A scatter plot of NSI against energy of 0.4-s sound segments with 0.2-s overlap ($R = 0.44$, $P < 0.01$). **b**, NSI (black line) and energy (purple line) as functions of time (same segment). **c**, A spectrogram of a non-separable, fluctuating segment (NSI = 0.45, FI = 28.6). **d**, A spectrogram of a separable, fluctuating segment (NSI = 0.08, FI = 28.0). The location of the segments in **c** and **d** are shown in **b** by arrowheads, and in **a** by circles. **e**, A scatter plot of NSI against FI for 0.4-s segments covering the whole data set. Only 64% of the segments appear in this plot; most of the others have larger NSI and/or FI values.

frequency alone and a function of time alone. We developed a non-separability index (NSI) to quantify the amount of discrepancy between the spectrogram and its separable approximation (the most similar separable function of time and frequency). In Fig. 1, the spectrogram is essentially equivalent to its approximation (NSI = 0.01; rather small NSI, see Fig. 2e). Segments with a gaussian carrier and separable spectrogram are approximately gaussian processes multiplied by a temporal envelope. We developed a fluctuation index (FI) to quantify the excess fluctuations of the temporal envelope relative to the expected amount for an unmodulated gaussian process with the same power spectrum. The sound segment in Fig. 1 has a significant excess of temporal fluctuations (FI = 17.1). Separable sounds with high excess temporal fluctuations have the important property that energy in different frequency bands is modulated coherently in time.

Many animal vocalizations contain frequency-modulated components, in which frequency and time are not separable. In contrast, mixtures of vocalizations, and non-animal sounds, tend to be separable. Figure 2a–d shows the analysis of a soundscape that contained vocalizations of a dominant bird over a background of other animals. The NSI fluctuates over time, and tends to be large when the total energy of the segment is large, corresponding to a call of the dominant bird. Thus, in this case, background sounds tend to be separable. To calibrate the scale of the NSI and FI, Fig. 2e shows their common distribution. The marginal distributions of NSI and FI have a peak near 0, consisting of those sounds that are separable (for the NSI) or have nonsignificant temporal fluctuation (for the FI). Based on these histograms, we considered segments with NSI < 0.1 to be separable and segments with FI < 5 to have nonsignificant envelope fluctuations. Approximately 29% of the

sounds are separable (NSI < 0.1), and also show larger than expected envelope fluctuations (FI > 5).

Mixtures of calls are probably separable because of a central limit theorem: when many calls are added together at random times, the result is a stationary gaussian process with the average power spectrum of the original calls. In simulations, five simultaneous calls were sufficient to reach the convergence criterion (NSI < 0.1). However, the energy fluctuations of different frequency bands became decorrelated and FI at convergence was small. Two possible factors may increase co-modulation. First, animal calls sometimes tend to form clusters. This effect was observed occasionally in the data set. Second, sounds acquire slow (<50 Hz), large envelope fluctuations during atmospheric propagation because of microturbulence⁹.

Thus, separable sounds with significant envelope fluctuations are rather common in the physical world. Because such separable backgrounds can elicit CMR, our results indicate that CMR is an adaptation of the human auditory system for detecting signals over naturally occurring separable backgrounds. We propose that other animals might also profit from the use of the same regularity in the structure of natural sounds. Indeed, it has been shown behaviourally that some birds show CMR¹⁰, although a physiological study in cochlear nucleus of cats did not find strong effects¹¹.

To test this hypothesis further, we searched for neuronal correlates of CMR in primary auditory cortex (AI) of cats (Figs 3, 4). In response to modulated noise bands (MNBs), most of the neurons (35/56, 63%) modulated their firing rates coherently with the temporal envelope (a phenomenon described previously¹² and referred to here as 'envelope locking') (Fig. 3a, modulated noise, red lines). In contrast, responses to unmodulated noise bands

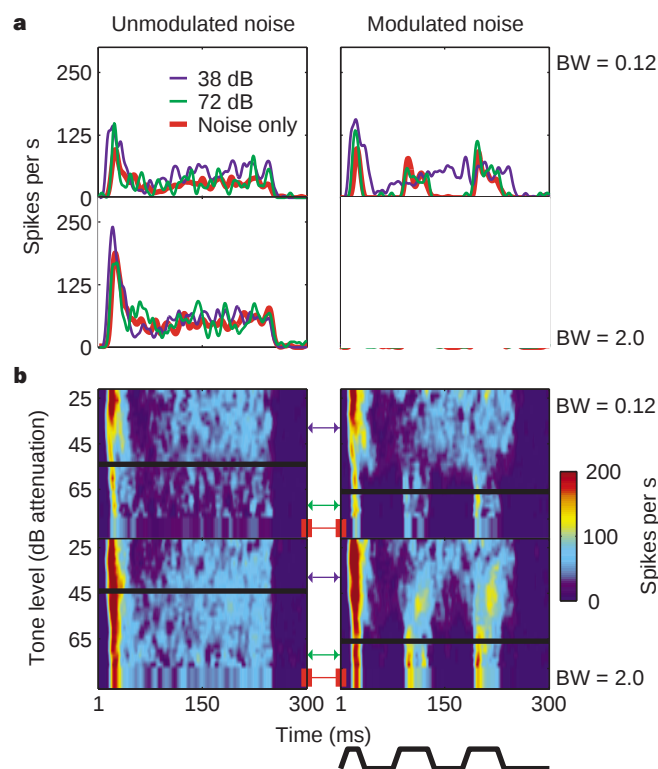


Figure 3 Responses to noise alone and to tone plus noise. **a**, Responses of a neuron to noise alone (red) and tone plus noise at two tone levels (purple and green). **b**, Responses to tone plus noise combinations at many levels. The tone levels used in **a** are marked by arrows (all levels are given in dB attenuation). The thick black lines are tone thresholds in noise. Thresholds: 54 dB (unmodulated), 66 dB (modulated) at 0.125 kHz; 44 dB (unmodulated), 66 dB (modulated) at 2 kHz. BW, bandwidth.

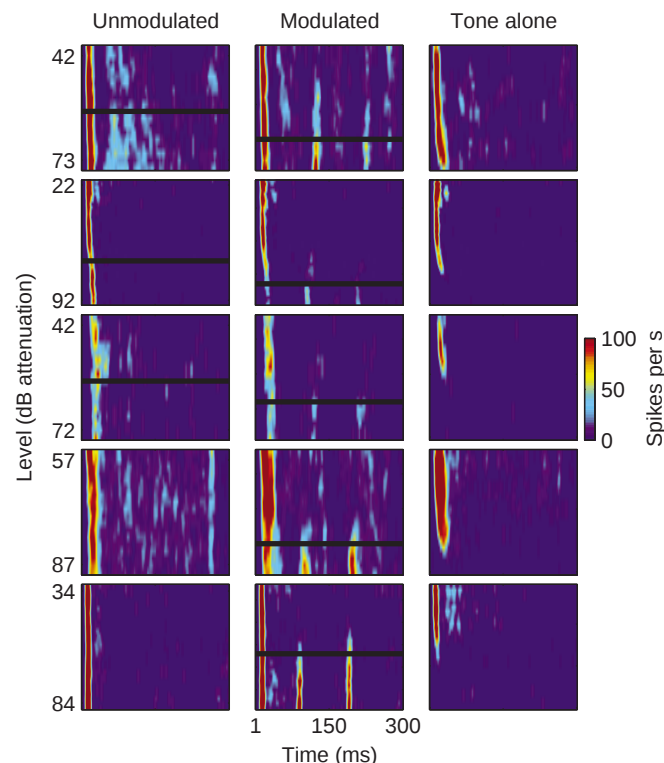


Figure 4 Additional examples of CMR. Each line shows the responses of one neuron to tone plus unmodulated noise (left), tone plus modulated noise (centre), and tone alone (right). Changes in envelope locking sometimes occur below the neuron's threshold in silence. From top to bottom: BF = 10.5 kHz, noise bandwidth (BW) = 10 kHz, threshold difference ($\Delta\theta$) = 8 dB; BF = 10.5 kHz, BW = 4.25 kHz, $\Delta\theta$ = 14 dB; BF = 3.4 kHz, BW = 1.7 kHz, $\Delta\theta$ = 6 dB; BF = 7.0 kHz, BW = 3.5 kHz, $\Delta\theta$ > 30 dB; BF = 7.8 kHz, BW = 3.9 kHz, $\Delta\theta$ > 34 dB. In the last two cases, threshold in unmodulated noise was not reached.

(UNBs) were either pure onset (23/35) or onset followed by unpatterned, sustained increases in firing rates (12/35; Fig. 3a, unmodulated noise, red lines).

We found that in a substantial fraction of the neurons that locked to the temporal envelope (25/35, 71%), the envelope locking was labile and degraded easily when a tone was added to the noise. In Fig. 3a, at the lower tone level (green lines), the response to the tone plus noise combination is essentially identical to the response to the noise alone. In contrast, at the higher tone level (blue lines), the temporal pattern of the response is modified. In responses to UNBs, sustained firing rates are increased, the onset response is larger and its latency is shorter. Similar changes in onset response also occur in responses to MNBs; however, in this case there is also a strong decrease in envelope locking, which is replaced by a more sustained firing pattern. Thus, if the shape of the peristimulus time histogram (PSTH) is used to distinguish between the noise alone and the tone plus noise cases, the detection threshold should occur between the two sound levels in all panels of Fig. 3a. To find the masked threshold, responses at many tone levels are presented as response planes, with firing rate displayed in colour as a function of tone level and time (Fig. 3b). The suppression of envelope locking caused a substantial reduction in detection thresholds of tones in MNBs compared with UNBs, and this reduction increased with bandwidth. Additional examples of the same phenomenon are shown in Fig. 4: neuronal responses lock to the envelope of the MNBs at low tone levels, at which the response to the tone plus noise combination is essentially identical to the responses to the noise alone (as in Fig. 3a at the low tone level). This envelope locking disappears at higher tone levels, giving a qualitative (as well as quantitative) signal for the presence of the added tone.

Envelope locking often increased with bandwidth when measured by locked rate, at least for the narrower bandwidths (23/35, 66%; see refs 13, 14). Thus, the population signal for release from masking in MNBs increased with bandwidth. Population modulated-unmodulated threshold differences, computed from the average response planes of the whole population reported here, were about 10 dB at the narrowest noise bandwidth tested (BF/32 of each neuron, where BF is the best frequency), and over 30 dB at the largest (BF/2 of each neuron). The decrease in threshold in MNBs as bandwidth was increased was over 20 dB in the population signal; using similar stimuli, corresponding values in humans are 10–15 dB (ref. 15). The decrease in thresholds as noise bandwidth is increased beyond the peripheral filters is the defining characteristic of true CMR. The degradation of envelope locking depended on time: later response components were often degraded more easily (for example, Fig. 4, top three examples); indeed, in human psychophysical experiments, sounds must be rather long (>200 ms) to elicit maximal CMR. Thus, neuronal responses and human psychophysics followed similar rules^{6,7}. With the cautionary note that the physiology of anaesthetized cats has been compared with the behaviour of humans, we speculate that this neuronal mechanism underlies CMR.

It is not clear whether the effects shown here are generated at the level of the primary auditory cortex. One argument for a cortical origin of these responses is the low modulation frequencies that are required to elicit CMR (above 50 Hz, true CMR is minimal). These rates correspond better to cortical modulation sensitivity than to modulation sensitivity of inferior colliculus neurons, which may follow amplitude modulation up to hundreds of Hz. The lability of the late envelope-following responses probably results from the sluggishness of cortical neurons.

We suggest that these results are a special case of a general correspondence principle for auditory physiology: regularities in the auditory biotope should be manifested in response properties of auditory neurons. Such correspondences have been suggested previously for species-specific vocalizations⁴; we have demonstrated here that this principle may cover a much larger class of auditory phenomena. □

Methods

Analysis of natural sounds. Recordings are the property of the Library of Natural Sounds, Cornell Laboratory of Ornithology, Ithaca, New York. The sounds were copied from digital audio tape to disk files. Sampling rate was reduced from 44.1 to 14.7 kHz, because the sounds did not contain appreciable energy above 7 MHz. Multiple 32-point fast Fourier transforms (FFTs) were computed, shifting the window by one sample between successive computations. The spectrogram was formed by taking the squared absolute value of the samples at the output of each FFT filter and smoothing them with a two-bin window in frequency. The carrier was computed by dividing the outputs of the FFT filters by the square root of the smoothed spectrogram; the result was transformed back into the time domain. Only one spectrum out of every 16 samples in time was kept. The frequency-dependent component was estimated by the average of the spectrogram over time, the temporal envelope by the average over frequency, and the separable approximation by their outer product.

The analysis of spectrotemporal separability was conducted on 0.4-s segments; this duration was chosen as it corresponds roughly to optimal durations for evoking CMR in humans. The analysis was also repeated for 1-s segments, without significant differences. The similarity between the spectrogram and the separable approximation was quantified using their mutual information. The expected value of the mutual information is positive even when the model is exact, because the carrier is noisy. Therefore, the expected value of the mutual information when the model is exact was computed using Monte-Carlo simulations and subtracted to form the NSI. The full range of NSI in the data set was between -0.3 and 0.7 bits; typical values were between -0.1 and 0.4.

Envelope fluctuations were quantified by the coefficient of variation of the temporal envelope. Because this measure is always positive, it was compared with its expected value for gaussian processes with the same average power spectrum but without temporal modulation, using Monte-Carlo simulations. The FI is their difference, expressed as a dimensionless *t*-value. The full range of FI in the data set was between -50 and 200; typical values were between -5 and 20.

Physiological recordings. Recordings were made in primary auditory cortex of five halothane-anaesthetized cats. Anaesthesia was induced by ketamine and xylazine and maintained with halothane (0.25–1.5%) using standard protocols authorized by the committee for animal care and ethics of the Hebrew University—Hadassah Medical School. Stimuli were generated digitally. Noise bands and tones were mixed after digital-to-analog conversion. Signals were presented to the animal through sealed, calibrated earphones (Sokolich). Frequency response was within ± 10 dB of its mean between 0.1–40 kHz; 0 dB attenuation is 110 dB SPL for tones and 60 dB Hz^{-1/2} for noise bands in the range of frequencies used here. To determine CMR, noise bands of varying width (between BF/32 and BF), which were centred arithmetically at BF, were used. Tones at BF were presented at various levels (from below to above masked threshold) in random order. For response to modulated noise, the noise band was multiplied by a 10-Hz sinusoidal or trapezoidal envelope. In all examples shown in this study, except for the top row in Fig. 4, the spectrum level of all the maskers was identical, causing a 3-dB difference in total energy between the modulated and unmodulated maskers. Single neurons were recorded using a glass-coated tungsten electrode and an online spike sorter (MSD, Alpha-Omega). All the neurons were well separated. To determine tone thresholds in noise, the variance-normalized squared differences between the noise and tone plus noise PSTHs were summed over the stimulus duration for each tone level; thresholds are levels above which these sum of squares increased significantly by a χ^2 test. This approach mimics to some extent the psychophysical paradigm of 2I2AFC in that we compare directly a 'noise interval', characterized by the responses to noise alone, with the 'tone plus noise' interval, characterized by the responses to the combined stimulus.

Received 5 August; accepted 5 November 1998.

1. Pelleg-Toiba, R. & Wollberg, Z. Discrimination of communication calls in the squirrel monkey: "Call detectors" or "cell ensembles"? *J. Basic Clin. Physiol. Pharmacol.* **2**, 257–272 (1991).
2. Steinschneider, M., Arezzo, J. C. & Vaughan, H. G. Jr Tonal features of speech-evoked activity in primate auditory cortex. *Brain Res.* **519**, 158–168 (1990).
3. Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* **74**, 2685–2706 (1995).

4. Suga, N. Philosophy and stimulus design for neuroethology of complex-sound processing. *Phil. Trans. R. Soc. Lond. B* **336**, 423–428 (1992).
5. Aertsen, A. M. H. J., Smolders, J. W. T. & Johannesma, P. I. M. Neural representation of the acoustic biotope: on the existence of stimulus-event relations for sensory neurons. *Biol. Cybern.* **32**, 175–185 (1979).
6. Hall, J. W., Haggard, M. P. & Fernandes, M. A. Detection in noise by spectro-temporal pattern analysis. *J. Acoust. Soc. Am.* **76**, 50–56 (1984).
7. Schooneveldt, G. P. & Moore, B. C. Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration. *J. Acoust. Soc. Am.* **85**, 273–281 (1989).
8. Ruderman, D. L. & Bialek, W. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* **73**, 814–817 (1994).
9. Richards, D. G. & Wiley, R. H. Reverberations and amplitude fluctuations in the propagation of sound in a forest: implication for animal communication. *Am. Nat.* **115**, 381–399 (1980).
10. Klump, G. M. & Langemann, U. Comodulation masking release in a songbird. *Hearing Res.* **87**, 157–164 (1995).
11. Rhode, W. S. & Greenwood, D. D. in *Abstracts of the 18th Association for Research in Otolaryngology Meeting* 127 (St Petersburg Beach, Florida, 1995).
12. Schreiner, C. E. & Urbas, J. V. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hearing Res.* **32**, 49–63 (1988).
13. Rauschecker, J. P., Tian, B. & Hauser, M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* **268**, 111–114 (1995).
14. Eggermont, J. J. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. *Hearing Res.* **74**, 51–66 (1994).
15. Carlyon, R. P., Buus, S. & Florentine, M. Comodulation masking release for three types of modulator as a function of modulation rate. *Hearing Res.* **42**, 37–45 (1989).

Acknowledgements. This work was supported by a grant administered by the Israel Science Foundation. We thank E. Vaadia, M. Abeles, E. Young and A. Aertsen for critical comments to this manuscript.

Correspondence and requests for materials should be addressed to I.N. (israel@music.md.huji.ac.il).

Kainate receptors mediate synaptic transmission between cones and ‘Off’ bipolar cells in a mammalian retina

Steven H. DeVries* & Eric A. Schwartz†

* Department of Ophthalmology and Visual Science, University of Texas Houston Health Science Center, Houston, Texas 77030, USA

† Department of Pharmacological and Physiological Sciences, University of Chicago, Chicago, Illinois 60637, USA

Light produces a graded hyperpolarization in retinal photoreceptors^{1,2} that decreases their release of synaptic neurotransmitter^{3,4}. Cone photoreceptors use glutamate^{5,6} as a neurotransmitter with which to communicate with two types of bipolar cell. Activation of metabotropic glutamate receptors in ‘On’ bipolar cells^{7,8} initiates a second-messenger cascade that can amplify small synaptic inputs from cones. In contrast, it is not known how the ionotropic glutamate receptors that are activated in ‘Off’ bipolar cells^{9,10} are optimized for transmitting small, graded signals. Here we show, by recording from a cone and a synaptically connected ‘Off’ bipolar cell in slices of retina from the ground squirrel, that transmission is mediated by glutamate receptors of the kainate-preferring subtype. In the dark, a cone releases sufficient neurotransmitter to desensitize most postsynaptic kainate receptors. The small postsynaptic current that persists (<5% of maximum) is quickly modulated by changes in presynaptic voltage. Since recovery from desensitization is slow (the decay time constant is roughly 500 milliseconds), little recovery can occur during the brief (roughly 100-millisecond) hyperpolarization that is produced in cones by a flash of light. By limiting the postsynaptic current, receptor desensitization prevents saturation of the ‘Off’ bipolar cell’s voltage response and allows the synapse to operate over the cone’s entire physiological voltage range.

We studied communication between cones and ‘Off’ bipolar cells in slices of the ground squirrel retina. ‘On’ and ‘Off’ bipolar cells were identified by both electrophysiological and anatomical criteria:

the two types of bipolar cell produce postsynaptic currents of opposite polarity and have axon terminals that end in different halves of the retina’s inner plexiform layer (IPL)¹¹. In the experiment illustrated in Fig. 1, the bipolar cell produced an inward postsynaptic current (Fig. 1c) and dye filled an axon terminal in the outer half (or sublamina a) of the IPL (Fig. 1a, b). Thus, both electrophysiology and anatomy showed that the recording was from an ‘Off’ bipolar cell. We studied 99 ‘Off’ bipolar cells which could be subdivided into at least three groups (types b2, b3 and b7 in ref. 12) on the basis of the pattern of their axon terminals in sublamina a. Nonetheless, the electrophysiological properties reported here were essentially the same in all cell pairs. We also observed bipolar cells that produced an outward current and had axon terminals in sublamina b. These were ‘On’ bipolar cells and are not discussed further.

We were immediately struck by the shape of the postsynaptic current (Fig. 1c). A large transient (average –228 pA; range –53 to –926 pA; $n = 99$) quickly declined (decay constant (τ) = 4.9 ± 2.9 ms) to a steady level (-9.5 ± 7.7 pA or $4.4 \pm 3.9\%$ of the maximum). We wanted to know how the transient and steady components might contribute to the reliable transmission of small, graded signals produced by a light-stimulated cone photoreceptor.

Our first results showed that Ca^{2+} -dependent transmitter release was activated throughout a cone’s physiological operating range (–60 to –35 mV) and was responsible for both components of the postsynaptic current. The membrane potential of a cone was stepped from –80 mV to a series of depolarized voltages, first in the presence of Ca^{2+} and then in a saline solution containing 1 mM Cd^{2+} (and no Ca^{2+}). Exposure to Cd^{2+} blocked the presynaptic Ca^{2+}

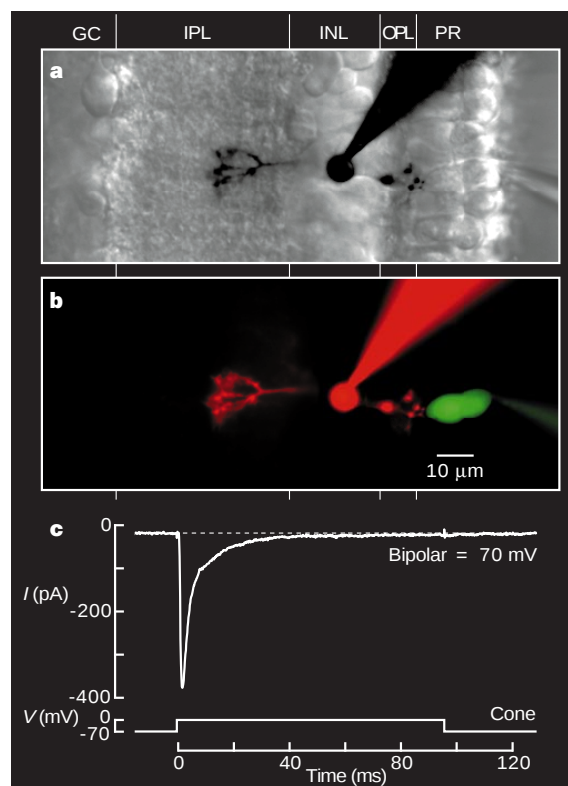


Figure 1 Synaptic transmission from cone photoreceptors to ‘Off’ bipolar cells. **a**, Sulphorhodamine 101 fluorescence (dark areas) in the bipolar cell is superimposed onto a Nomarski micrograph of the retinal slice. **b**, Sulphorhodamine 101 fluorescence (red) in the bipolar cell and BODIPY 492/515 fluorescence (green) in the cone are superimposed. **c**, Postsynaptic response produced when the cone’s membrane voltage was stepped from –70 to 0 mV. Bipolar cells were held at –70 mV. PR, photoreceptors; OPL, outer plexiform layer; INL, inner nuclear layer; GC, ganglion cells.

current (Fig. 2a) and the postsynaptic current (Fig. 2c, d, seven pairs). The Ca^{2+} current began at -60 mV, increased as the voltage approached -40 mV, and then declined with steps to more depolarized voltages beyond the physiological range (Fig. 2b). Both peak and steady postsynaptic currents increased and declined with the amplitude of the presynaptic Ca^{2+} current (Fig. 2b, d–f).

A rapid decline from an initial peak might be produced either by a decrease in the rate of vesicle release¹³ or by a desensitization of postsynaptic receptors. Information about both mechanisms can be obtained by measuring the time course of return to full sensitivity. For this purpose, we measured the interaction between two identical depolarizing pulses. The first pulse produced a maximum synaptic response; a second pulse delivered a short time later produced a smaller response (Fig. 3a). The time course for recovery of the maximum amplitude is plotted as a function of interpulse interval in Fig. 3c (circles). The time constant for recovery was 481 ± 127 ms ($n = 6$).

The following results show that desensitization of postsynaptic receptors is sufficient to explain the slow synaptic recovery. Ionotropic glutamate receptors can be divided into three families¹⁴: AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid), kainate and NMDA (*N*-methyl-D-aspartic acid) receptors. A selective inhibitor of NMDA receptors, amino-phosphonovaleric acid, had no effect on synaptic transmission (data not shown). In addition, the current–voltage relationship for the postsynaptic current was linear between -50 and $+40$ mV (data not shown) and lacked the characteristic rectification normally seen when NMDA receptors are bathed in a saline solution containing Mg^{2+} (ref. 15). Hence, NMDA receptors did not contribute to the synaptic response. In contrast, an inhibitor of both AMPA and kainate receptors, 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) reversibly inhibited synaptic transmission (Fig. 4a, b; $n = 12$). The distinction between AMPA and kainate receptors was resolved by using GYKI 53655, a selective, non-competitive blocker of AMPA receptors¹⁶. GYKI 53655 (10 – 25 μM) did not affect synaptic transmission (Fig. 4c); the ratio of peak current amplitudes observed in saline solution in the presence and absence of GYKI 53655 was 1.00 ± 0.07 ($n = 7$). In addition, transmission was not affected when two cell pairs were superfused with cyclothiazide (200 μM), an agent that prolongs the activation of AMPA receptors¹⁷. Both

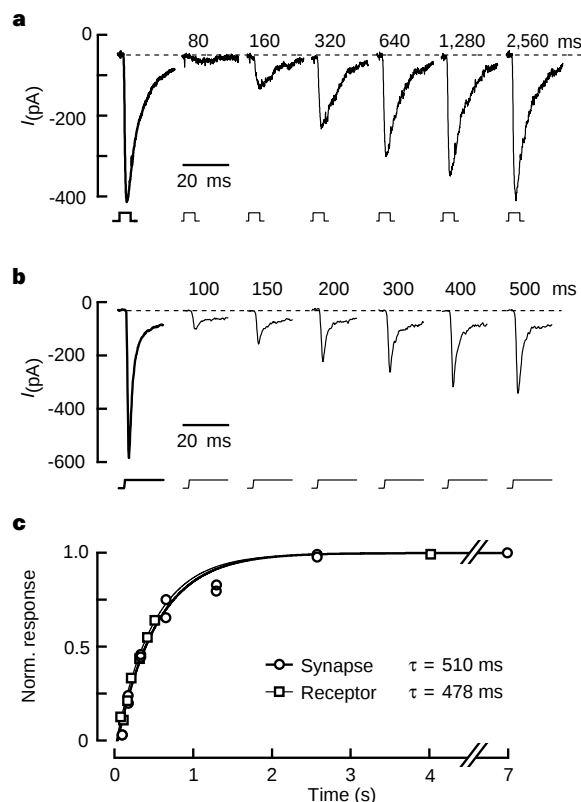


Figure 3 Rate of recovery of synaptic and kainate-receptor sensitivity. **a**, Paired depolarizing pulses were delivered to a cone (5-ms duration for each pulse; 7 s between pairs) and the postsynaptic current in an 'Off' bipolar cell was recorded. **b**, Paired pulses of 1 mM glutamate were applied to an isolated 'Off' bipolar cell (4 s between pairs, 25 μM GYKI 53655 present at all times). In **a**, **b**, the average response to the first pulse is at the left; the responses to the second pulses are shown at the right, with the interpulse interval given above each trace. **c**, Normalized peak response measured from the data in **a** (circles) and **b** (squares) is plotted as a function of interpulse interval. Each line is a least-squares fit of a single exponential function.

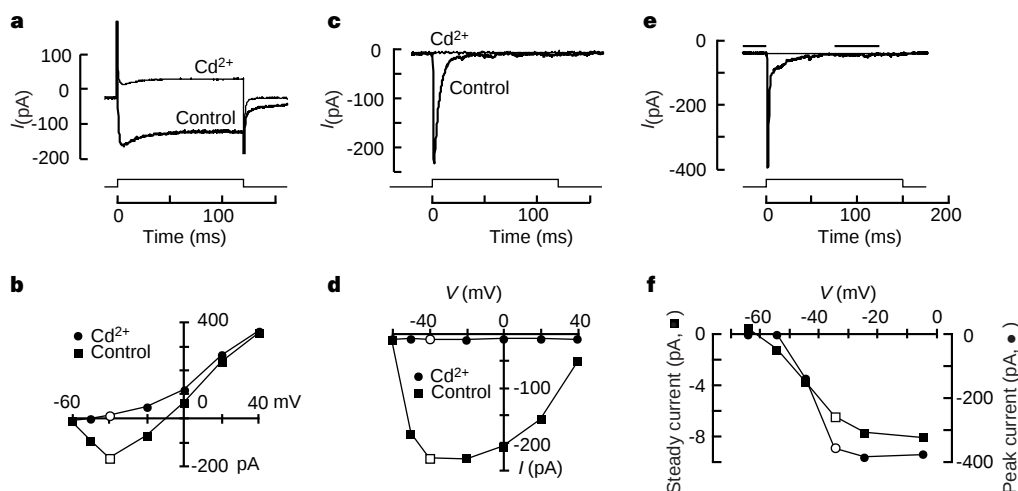


Figure 2 Ca^{2+} and voltage dependence of synaptic transmission. **a**, A whole-cell current was recorded in a cone while the slice was superfused first with normal saline and then with a saline solution containing Cd^{2+} . The voltage of the cone was stepped from -80 to -40 mV (as indicated by the timing trace, bottom). **b**, Current–voltage plot for a set of presynaptic voltage steps measured as in **a**. **c**, Whole-cell current of an 'Off' bipolar cell (holding potential -70 mV) recorded simultaneously with the records shown in **a**. **d**, Plot of peak postsynaptic current against

presynaptic voltage. **e**, Postsynaptic response recorded when the voltage of a cone was stepped from -80 to -35 mV (as indicated by the timing trace, bottom). **f**, Peak and steady postsynaptic currents plotted against presynaptic voltage. Steady currents were determined by averaging the current in each trace during the times indicated by the bars above the trace in **e** and taking the difference. **a–d** show results from one pair of cells and **e**, **f** show results from another pair. Open symbols in **b**, **d**, and **f** were measured from corresponding traces in **a**, **c**, and **e**.

GYKI 53655 and cyclothiazide were effective at AMPA receptors expressed by amacrine cells in the same slices (data not shown). These results indicate that kainate receptors may mediate synaptic transmission between cones and 'Off' bipolar cells.

We studied the properties of kainate receptors by using individual 'Off' bipolar cells isolated from retinal slices (Fig. 4d). Kainate itself can discriminate between kainate and AMPA receptors: it activates both types of receptor, but the current through kainate receptors is quickly curtailed by desensitization whereas the current through AMPA receptors is sustained¹⁸. The direct application of kainate to an isolated bipolar cell produced a large transient and a smaller steady current. The response was only slightly smaller when the application of kainate was repeated in the presence of 25 μ M GYKI 53655 (Fig. 4e). As expected (Fig. 4f), glutamate (in the presence of 25 μ M GYKI 53655) also produced a response with a large transient (-657 ± 312 pA; $n = 4$) that declined ($\tau = 1.9 \pm 0.4$ ms) to a smaller steady level ($2.2 \pm 1.0\%$ of maximum). When the glutamate pulse ended, the steady response declined with a deactivation time constant of 5.5 ± 0.8 ms. However, the recovery of full sensitivity was much slower. We measured the time course of the return to the fully sensitized state by delivering a delayed, second pulse of glutamate (Fig. 3b). The normalized peak amplitude of the second response is plotted as a function of interpulse interval in Fig. 3c (squares). The time constant for recovery was 480 ms (range 480–682 ms; $n = 3$), remarkably similar to the time course for recovery of the synaptic response.

Desensitization will shape the synaptic response if most of the postsynaptic receptors are activated during the transient. The peak synaptic current produced by activating a single cone (average = -228 pA) was a significant fraction of the maximum current produced when glutamate was applied to an isolated cell (average = -657 pA). Thus, it seems likely that a large depolarizing step in a cone (see, for example, Fig. 1c) produced a surge of transmitter release that activated most or all of the postsynaptic receptors. Thereafter, the concentration of transmitter in the synaptic cleft probably decreased but remained sufficient to desensitize most receptors.

The effect of desensitization was evident when we compared the

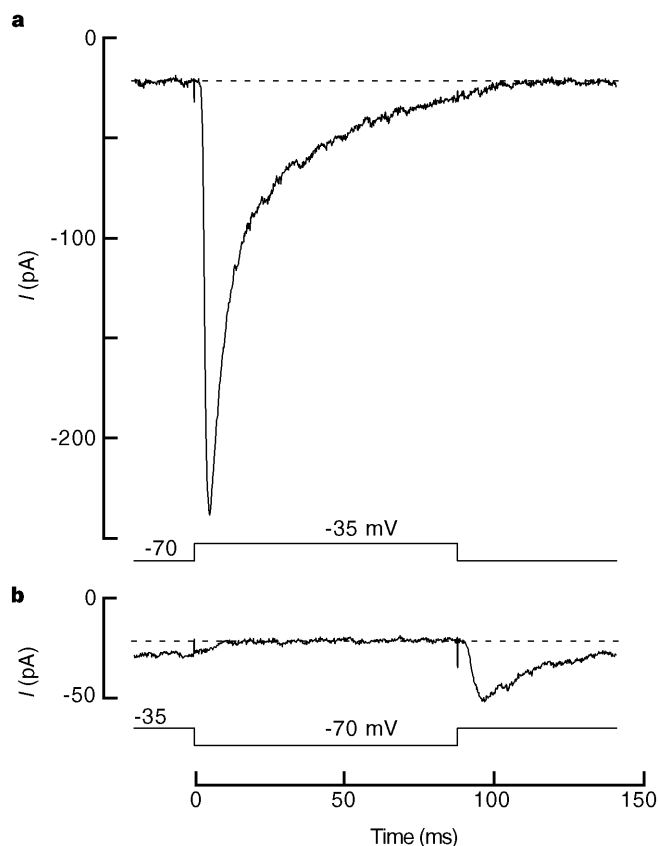


Figure 5 The postsynaptic current is reduced by receptor desensitization at physiological voltages. **a**, The membrane voltage of a cone was stepped from -70 to -35 mV for 90 ms and the postsynaptic current was recorded. **b**, The membrane voltage of the same cone was stepped from -35 to -70 mV.

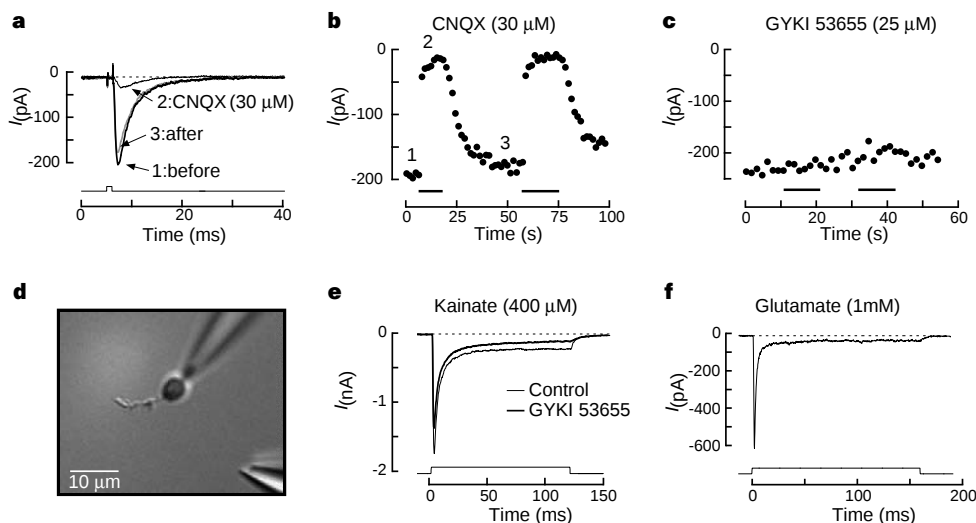


Figure 4 Pharmacological characterization of synaptic receptors. **a**, Postsynaptic responses recorded 1, before, 2, during and 3, after a slice was superfused with a saline containing 30 μ M CNQX. **b**, Peak synaptic current is plotted against time for the experiment shown in **a**. CNQX was applied during the intervals shown by the bars. Numbers refer to the traces in **a**. **c**, A similar procedure was used to test the

effect of 25 μ M GYKI 53655. **d**, A bipolar cell soma with attached dendrites was removed from a slice. A puffer pipette is shown at the lower right. **e**, The response of an isolated 'Off' bipolar cell to the rapid application of 400 μ M kainate in control saline and in saline that contained 25 μ M GYKI 53655. **f**, The response to 1.0 mM glutamate in 25 μ M GYKI 53655. Dashed line indicates holding current.

amplitude of an evoked miniature synaptic current (MSC) with the noise of synaptic communication. Synaptic noise is produced by both the random arrival of MSCs and the fluctuation in the number of open channels. We measured the noise produced by channels when a long step of glutamate was applied to a solitary bipolar cell. Dividing the increase in variance, $\Delta\sigma^2$, by the mean, I gave an estimate for the amplitude of a single, open, glutamate-gated channel, i , of -0.37 pA (two cells). For comparison, continuously depolarizing a cone produced a steady synaptic current of -6.8 pA (nine synapses) with a variance of 3.74 ± 3.60 (pA)². An estimate of channel noise is approximately $iI = -0.37$ pA $\times$ -6.8 pA = 2.4 (pA)²; thus the remaining 1.3 (pA)² is attributed to the stochastic arrival of MSCs. The latter component can be compared with a mean and variance calculated from the size and shape of an MSC by using Campbell's theorem^{19,20}. A threshold stimulus produced all-or-nothing evoked MSCs with a mean amplitude of -5 pA and a total charge of 20 fC (two synapses). Consequently, the mean steady current could be produced by MSCs of the observed size if vesicles were released at a rate of 340 per second. However, the same MSC size and rate of vesicle release gives a variance of ~ 20 (pA)², at least tenfold larger than the component attributed to the random arrival of MSCs. The discrepancy can be removed and both the observed mean and the variance can be reproduced if desensitization were to decrease the size of MSCs roughly tenfold, the rate of release being increased by a similar factor. This high rate of continuous exocytosis requires that synaptic vesicles be primed and readied for release at least tenfold faster than observed in goldfish bipolar cells²¹.

Receptor desensitization has a profound effect on the normal function of 'Off' bipolar cells. The familiar large transient and small steady current (Fig. 5a, voltage step from -70 to -35 mV) was replaced with a very different response when we mimicked the physiological change produced by a flash of light^{22,23} (Fig. 5b, voltage step from -35 to -70 mV). Prolonged depolarization produced a steady current of -7 pA that was quickly stopped by hyperpolarization (Fig. 5b). Subsequent depolarization produced a transient of -29 pA, an amplitude that would be predicted if desensitization in this bipolar cell recovered with a time constant of 590 ms during the 90 -ms step. A bipolar cell in the periphery of the ground squirrel retina sums the unequal input from 5 – 15 cones²⁴. If the total input were three- to fivefold larger than the current seen in Fig. 5, the combined input during darkness would be just enough to polarize a bipolar cell with a 1 -G Ω input resistance through its 25 -mV operating range. Because of desensitization of the bipolar-cell kainate receptors, the cone can release transmitter at a high rate and yet the bipolar cell can avoid a large conductance that would otherwise drive its membrane voltage to saturation and reduce its sensitivity to incremental changes in transmitter concentration²⁵. □

Methods

Retinal slices. Thirteen lined ground squirrels (*Spermophilus tridecemlineatus*) were killed by intracardiac injection of pentobarbital (20 mg). The eyes were removed and dissected. Pieces of the superior retina were placed vitreous side down onto a piece of filter paper (HVLP, Millipore) and the pigment epithelium was removed. Slices of retina and attached filter paper were cut to a thickness of ~ 100 μ m (ref. 26) and transferred in HEPES-buffered (10 mM, pH 7.35) AMES solution to a recording chamber. The tissue was continuously superfused with extracellular saline that contained (in mM): 115 NaCl, 24 NaHCO₃, 3.1 KCl, 2 CaCl₂, 1.24 MgSO₄, 6 glucose, 1 sodium pyruvate, 1 sodium lactate, 1 sodium malate, 1 sodium succinate, 0.1 BaCl₂, and 0.05% phenol red, equilibrated with 5% carbon dioxide to give a pH of 7.4 . The extracellular solution also contained 50 μ M of each of picrotoxin, strychnine and mecamylamine, and 0.4 μ M tetrodotoxin. In the Ca²⁺-free solution, 1 mM Cd²⁺ and 1 mM Mg²⁺ were substituted for Ca²⁺. Except as noted, the membrane voltages of both a cone and a hyperpolarizing bipolar cell were controlled through patch pipettes in the whole-cell, tight-seal configuration. The patch-pipette solution contained (in mM): 80 potassium aspartate, 20 caesium

aspartate, 10 CsCl, 2 MgCl₂, 1 CaCl₂, 10 EGTA, 1 ATP, 0.1 GTP, and 0.1 leupeptin. Sulphorhodamine 101 (0.5 mM) was included in the pipette attached to a bipolar cell and BODIPY 492/515 (0.5 mM; Molecular Probes) was included in the pipette attached to a cone. Membrane currents were low-pass-filtered at 4.7 kHz. Drugs were applied from a nearby puffer pipette. Slices were maintained at 32 – 33 °C during recording. Results are given as means ± 1 s.d.

In Figs 3a and 4a–c the pipette attached to the cone was left unpolished and a loose-seal was formed. Although intracellular access was not established the cone could still be polarized, as indicated by the bipolar-cell response. The loose-seal technique was useful for prolonged experiments involving the application of pharmacological agents or twin pulses to measure recovery of sensitivity, as it reduced the likelihood that synapse 'run-down' would occur.

Isolated bipolar cells. The axon of the targeted cell was first mechanically severed, and then a solution containing ~ 11 U ml⁻¹ papain (Worthington) dissolved in HEPES-buffered AMES medium was applied from a puffer pipette to the outer plexiform layer near the soma for 5 – 10 s. The recording pipette was then sealed to a bipolar cell. After a whole-cell recording was established, the recording pipette was used to gently remove the soma and dendrite from the slice. A rapid-perfusion system used either two- or four-barrelled pipettes mounted on a piezoelectric stepper motor. The temperature of saline in the rapid-perfusion pipette was not heated. The actual temperature at the cell may have been a few degrees less than the 32 – 33 °C maintained in the bath. Glutamate or kainate produced large inward currents in isolated 'Off' bipolar cells and little current in presumed 'On' bipolar cells.

Received 8 October; accepted 23 November 1998.

- Penn, R. D. & Hagins, W. A. Signal transmission along retinal rods and the origin of the electroretinographic *a*-wave. *Nature* **223**, 201–205 (1969).
- Baylor, D. A. & Fuortes, M. G. F. Electrical responses of single cones in the retina of the turtle. *J. Physiol. (Lond.)* **207**, 77–92 (1970).
- Trifonov, Y. A. Study of synaptic transmission between the photoreceptor and the horizontal cell using electrical stimulation of the retina. *Biofizika* **13**, 809–817 (1968).
- Dowling, J. E. & Ripps, H. Effect of magnesium on horizontal cell activity in the skate retina. *Nature* **242**, 101–103 (1973).
- Miller, A. M. & Schwartz, E. A. Evidence for the identification of synaptic transmitters released by photoreceptors of the toad retina. *J. Physiol. (Lond.)* **334**, 325–349 (1983).
- Copenhagen, D. R. & Jahr, C. E. Release of endogenous excitatory amino acids from turtle photoreceptors. *Nature* **341**, 536–539 (1989).
- Nawy, S. & Jahr, C. E. cGMP-gated conductance in retinal bipolar cells is suppressed by the photoreceptor transmitter. *Neuron* **7**, 677–683 (1991).
- Masu, M. *et al.* Specific deficit of the ON response in visual transmission by targeted disruption of the mGluR6 gene. *Cell* **80**, 757–765 (1995).
- Saito, T. & Kaneko, A. Ionic mechanisms underlying the responses of off-center bipolar cells in the carp retina. *J. Gen. Physiol.* **81**, 589–601 (1983).
- Attwell, D., Mobbs, P., Tessier-Lavigne, M. & Wilson, M. Neurotransmitter-induced currents in retinal bipolar cells of the axolotl, *Ambystoma mexicanum*. *J. Physiol. (Lond.)* **387**, 125–161 (1987).
- Famiglietti, E. V. & Kolb, H. Structural basis for ON- and OFF-center responses in retinal ganglion cells. *Science* **194**, 193–195 (1976).
- West, R. W. Light and electron microscopy of the ground squirrel retina: functional considerations. *J. Comp. Neurol.* **168**, 355–378 (1976).
- von Gersdorff, H. & Matthews, G. Dynamics of synaptic vesicle fusion and membrane retrieval in synaptic terminals. *Nature* **367**, 735–739 (1994).
- Ozawa, S., Kamiya, H. & Tsuzuki, K. Glutamate receptors in the mammalian central nervous system. *Prog. Neurobiol.* **54**, 581–618 (1998).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. Voltage-dependent block by Mg²⁺ of NMDA responses in spinal cord neurones. *Nature* **309**, 261–263 (1984).
- Patrain, A. V., Morales, M. & Lerma, J. Selective antagonism of AMPA receptors unmasks kainate receptor-mediated responses in hippocampal neurons. *Neuron* **14**, 185–189 (1995).
- Partin, K. M. *et al.* Selective modulation of desensitization at AMPA versus kainate receptors by cyclothiazide and Concanavalin A. *Neuron* **11**, 1069–1082 (1993).
- Patneau, D. K. *et al.* Glial cells of the oligodendrocyte lineage express both kainate- and AMPA-preferring subtypes of glutamate receptor. *Neuron* **12**, 357–371 (1994).
- Rice, S. O. Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**, 282–332 (1944).
- Katz, B. & Miledi, R. The statistical nature of the acetylcholine potential and its molecular components. *J. Physiol. (Lond.)* **224**, 665–699 (1972).
- Mennerick, S. & Matthews, G. Ultrafast exocytosis elicited by calcium current in synaptic terminals of retinal bipolar neurons. *Neuron* **17**, 1241–1249 (1996).
- Leeper, H. F. & Charlton, J. S. Response properties of horizontal cells and photoreceptor cells in the retina of the tree squirrel, *Sciurus carolinensis*. *J. Neurophysiol.* **54**, 1157–1166 (1985).
- Kraft, T. Photocurrents of cone photoreceptors of the golden-mantled ground squirrel. *J. Physiol. (Lond.)* **404**, 199–213 (1988).
- Lindberg, K. A., Suemune, S. & Fisher, S. K. Retinal neurons of the California ground squirrel, *Spermophilus beecheyi*: a Golgi study. *J. Comp. Neurol.* **365**, 173–216 (1996).
- Falk, G. Signal transmission from rods to bipolar and horizontal cells: a synthesis. *Prog. Retinal Res.* **8**, 255–279 (1989).
- Lukasiewicz, P. & Werblin, F. A slowly inactivating potassium current truncates spike activity in ganglion cells of the tiger salamander retina. *J. Neurosci.* **8**, 4470–4481 (1988).

Acknowledgements. This work was supported by an NIH grant (to E.A.S.) and a Research to Prevent Blindness Junior Investigator Award (to S.H.D.).

Correspondence and requests for materials should be addressed to S.H.D. (e-mail: sdevries@eye.med.uth.tmc.edu).

Kainate-receptor-mediated sensory synaptic transmission in mammalian spinal cord

Ping Li*, Timothy J. Wilding†, Susan J. Kim*, Amelita A. Calejesan*, James E. Huettner† & Min Zhuo*‡

Departments of * Anesthesiology, † Cell Biology and Physiology, and ‡ Anatomy and Neurobiology, Washington University in St Louis, St Louis, Missouri 63110, USA

Glutamate, the major excitatory neurotransmitter in the central nervous system, activates three different receptors that directly gate ion channels, namely receptors for AMPA (α -amino-3-hydroxy-5-methyl isoxazole propionic acid), NMDA (*N*-methyl-D-aspartate), and kainate, a structural analogue of glutamate. The contribution of AMPA and NMDA receptors to synaptic transmission and plasticity is well established¹. Recent work on the physiological function of kainate receptors has focused on the hippocampus², where repetitive activation of the mossy-fibre pathway generates a slow, kainate-receptor-mediated excitatory postsynaptic current (EPSC)^{3–5}. Here we show that high-intensity

single-shock stimulation (of duration 200 microseconds) of primary afferent sensory fibres produces a fast, kainate-receptor-mediated EPSC in the superficial dorsal horn of the spinal cord. Activation of low-threshold afferent fibres generates typical AMPA-receptor-mediated EPSCs only, indicating that kainate receptors may be restricted to synapses formed by high-threshold nociceptive (pain-sensing) and thermoreceptive primary afferent fibres. Consistent with this possibility, kainate-receptor-mediated EPSCs are blocked by the analgesic μ -opiate-receptor agonist Damgo and spinal blockade of both kainate and AMPA receptors produces antinociception. Thus, spinal kainate receptors contribute to transmission of somatosensory inputs from the periphery to the brain.

Glutamate is the fast neurotransmitter between primary afferent fibres and dorsal horn sensory neurons⁶. Previous studies have shown that there are two components of transmission at these synapses: a rapid component blocked by the non-NMDA-receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and a slower component mediated by D(-)-2-amino-5-phosphonovaleric acid (AP5)-sensitive NMDA receptors⁷. Because CNQX blocks both AMPA and kainate receptors, this earlier work did not resolve the different contributions of these two receptor subtypes. To address this question, we obtained whole-cell patch-clamp recordings from neurons in the superficial dorsal horn of the spinal cord, which receive afferent inputs from primary sensory neurons in the

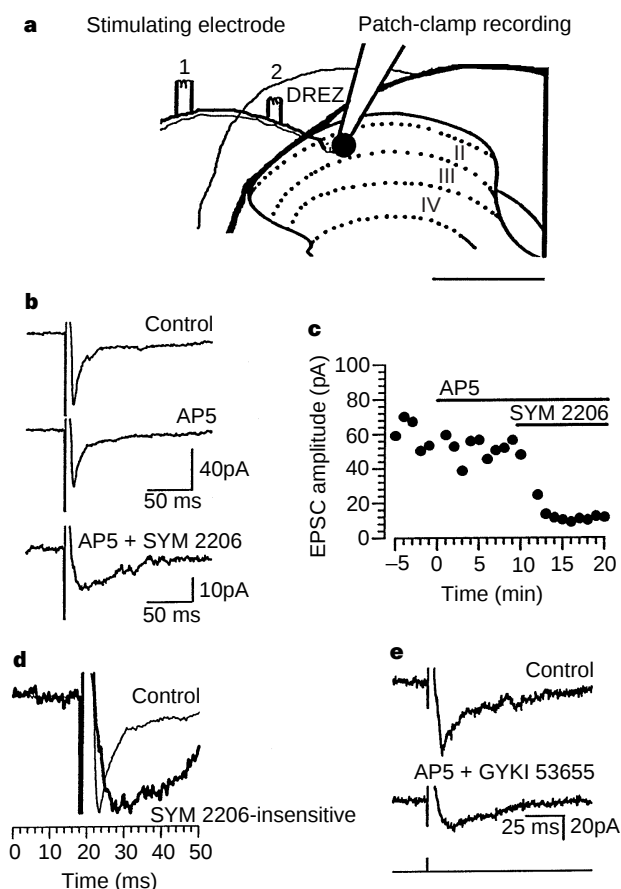


Figure 1 AMPA- and NMDA-receptor antagonists produce incomplete block of primary afferent EPSCs. **a**, Diagram of a spinal slice showing the placement of recording and stimulating (1, dorsal root nerves; 2, DREZ) electrodes. II, III and IV indicate laminae II, III and IV in the dorsal horn. Scale bar represents 1 mm. **b**, EPSCs recorded in control medium, 10 min after addition of AP5 (100 μ M), and 10 min after addition AP5 and SYM 2206. **c**, Peak EPSC amplitude versus time for the traces shown in **b**; **d**, The SYM-2206-insensitive EPSC scaled to the peak of the control EPSC. **e**, EPSCs recorded in control medium and 10 min after addition of AP5 (100 μ M) and GYKI 53655.

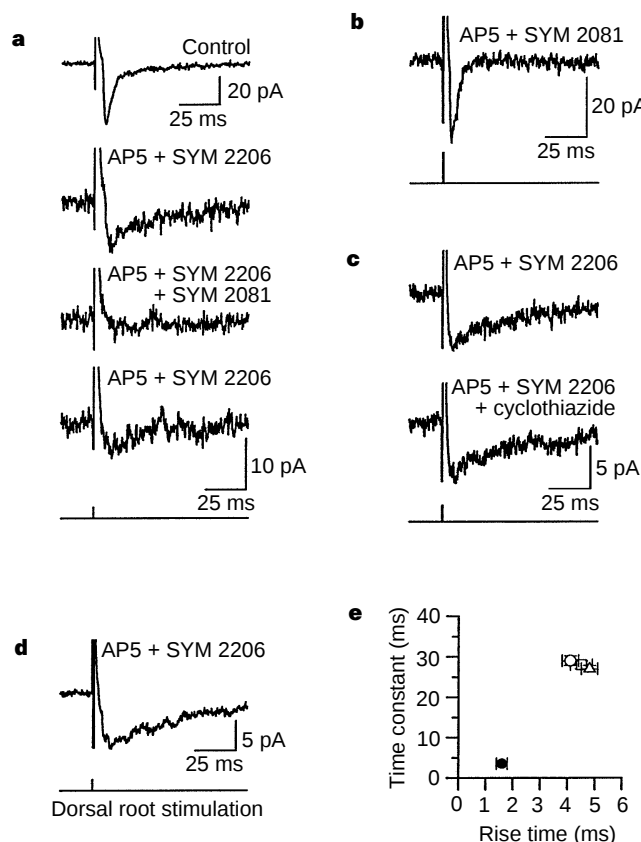


Figure 2 Kainate-receptor-mediated synaptic transmission. **a**, EPSCs in saline solution, in AP5 (100 μ M) and SYM 2206 (100 μ M) and AP5, SYM 2206 and SYM 2081 (1 μ M), and partial recovery in AP5 and SYM 2206. **b**, EPSCs in AP5 and SYM 2081 (5 μ M). **c**, SYM-2206-insensitive EPSCs were not affected by cyclothiazide (10 μ M). **d**, An EPSC produced by dorsal root stimulation in the presence of AP5 and SYM 2206. **e**, Time constant of EPSC decay versus the rise time (10–90%) for EPSCs in GYKI 53655 (open circle) or SYM 2206 (triangle), with dorsal root stimulation in SYM 2206 (square) and minimal stimulation in saline solution (filled circle).

periphery^{8,9}. Fast, monosynaptic EPSCs were induced by electrical stimulation delivered to the dorsal root entry zone (DREZ; Fig. 1a). We blocked NMDA receptors with the selective inhibitor AP5 (100 μ M) and tested the action of two selective, non-competitive AMPA-receptor antagonists, GYKI 53655 and SYM 2206 (refs 10–13). Each drug was used at 100 μ M, a concentration that produces maximal inhibition of AMPA receptors (half-maximal inhibitory concentration = 1–2 μ M) but less than 20–30% inhibition of kainate receptors^{10–13}.

Superfusion of slices with AP5 plus GYKI 53655 (Fig. 1e; $n = 10$) or AP5 plus SYM 2206 (Fig. 1b–d; $n = 10$) reduced, but did not completely block, the EPSCs. In contrast, AP5 plus 10 μ M CNQX or 2,3-dihydroxy-6-nitro-7-sulphamoylbenzo quinoxaline (NBQX) always produced almost complete inhibition of EPSCs ($n = 8$). For EPSCs studied before and after exposure to AMPA-receptor antagonists, the relative peak amplitude of the residual EPSC (the EPSC remaining after exposure to AMPA-receptor antagonists) was $31.3 \pm 5.6\%$ of control ($n = 20$ cells; 52.7 $\pm$ 7.0 pA before and 11.6 $\pm$ 1.7 pA after; 3 traces per cell for each condition). In two cases, the peak currents after superfusion of GYKI 53655 ($n = 1$) or SYM 2206 ($n = 1$) were small (2.6 pA and 3.8 pA). The rise time (10–90%) and decay time constant (τ) for residual EPSCs were significantly longer than for AMPA-receptor-mediated currents (Fig. 2e). To confirm that the residual currents were mediated by kainate receptors, we used SYM 2081, a selective kainate-receptor-agonist that causes potent receptor desensitization^{14,15}. SYM 2081 (1

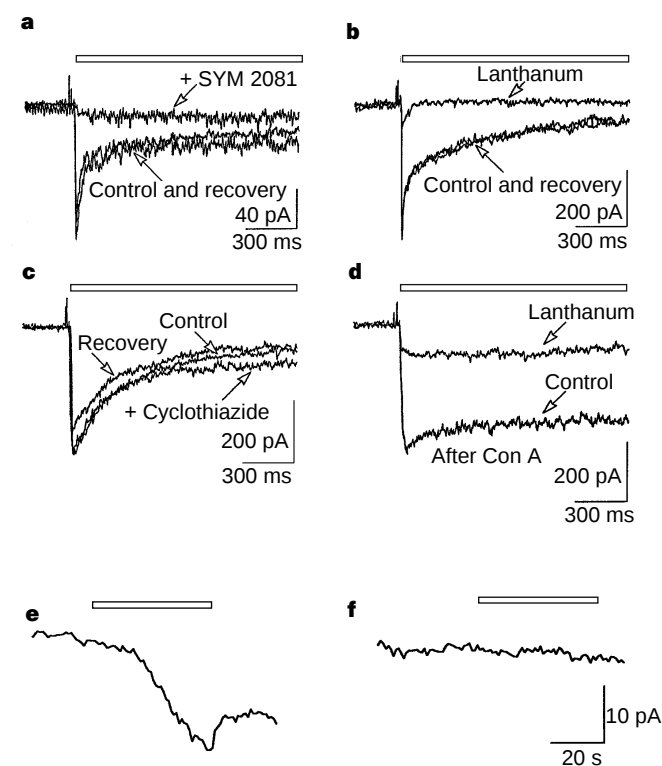


Figure 3 Kainate-receptor-mediated currents in spinal cultured neurons and a slice. **a–d**, Spinal cultured neurons; **e, f**, spinal slice. In **a–d**, GYKI 53655 was continuously superfused. **a**, Pre-exposure to 1 μ M SYM 2081 inhibited the current evoked by 300 μ M kainate (open bar; 76 $\pm$ 6% inhibition; $n = 5$). **b**, Kainate-evoked currents were inhibited by 15 μ M lanthanum (77 $\pm$ 11% inhibition; $n = 3$). **c**, 5 μ M cyclothiazide had little effect on kainate-evoked currents (101 $\pm$ 5% of control; $n = 4$). **d**, Receptor desensitization was reduced by pretreatment with conA ($n = 3$), but kainate-evoked current was still sensitive to inhibition by lanthanum. **e**, The current evoked by application of 10 μ M kainate to a spinal slice in 100 μ M SYM 2206, 100 μ M AP5 and 2 μ M Con A ($n = 8$). **f**, In the additional presence of 20 μ M CNQX, the current produced in **e** was not observed ($n = 5$).

or 5 μ M) significantly decreased or abolished the residual EPSCs in the presence of AP5 and SYM 2206 ($n = 3$; Fig. 2a), but did not affect the fast component of control EPSCs recorded in AP5 alone ($n = 4$; Fig. 2b). Cyclothiazide (10 μ M), which potentiates currents mediated by AMPA receptors but not by kainate receptors¹⁶, did not affect SYM-2206-insensitive EPSCs ($n = 5$; Fig. 2c). These results indicate that the residual EPSCs were mediated by kainate receptors.

Because stimulation of the DREZ might activate fibres from spinal glutamatergic interneurons and/or descending pathways¹⁷, in addition to primary afferent fibres, we performed experiments in which the stimulating electrode was positioned on an attached dorsal root nerve (Fig. 1a). In all cases, this method evoked EPSCs with a component that was resistant to block by SYM 2206 (Fig. 2d; $n = 6$).

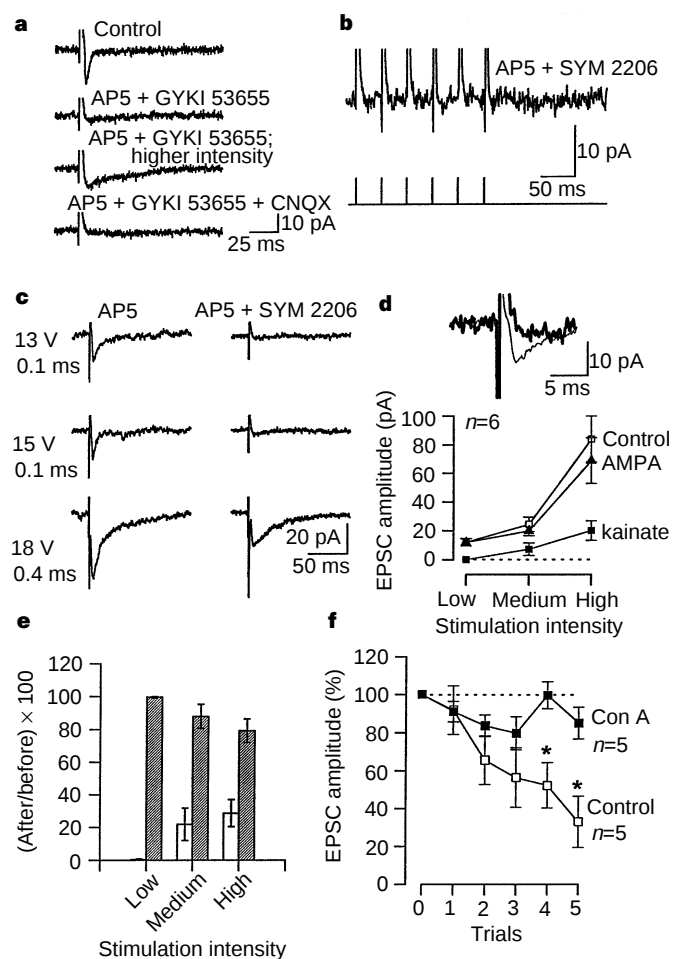


Figure 4 Kainate receptors reside at synapses receiving high-threshold afferents. **a**, Stimulation at low intensities induced GYKI-53655-sensitive AMPA currents (top two traces). At a higher intensity, a GYKI-53655-resistant current was induced and could be blocked by 10 μ M CNQX (bottom two traces). **b**, Repetitive stimulation in the presence of AP5 and SYM 2206 did not elicit kainate currents. **c**, Single traces of EPSCs induced by stimulation at different intensities in 100 μ M AP5 or AP5 plus 100 μ M SYM 2206. **d**, Bottom, summary of within-cell data from **c** (AP5, $n = 6$, open squares; AP5 plus SYM 2206, $n = 6$, filled squares). Inset, the control EPSC at low-stimulation intensity, scaled to 30% of its original amplitude (thin line) and the EPSC in SYM 2206 (thick line); single traces are shown. **e**, SYM-2206-sensitive (AMPA; shaded bars) and -resistant (kainate; open bars) EPSCs as a percentage of control EPSCs. Differences between the means within each set are significant ($P < 0.01$). Responses to low intensity are different from responses to medium and high intensity by pairwise comparison ($P < 0.05$). AMPA component is the control EPSC–kainate component. **f**, Pretreatment with conA (2 μ M) reduced the decay of kainate currents during 1-Hz stimulation. Asterisk indicates $P < 0.05$.

Although the existence of kainate receptors in the spinal cord has been reported^{18–20}, their subunit composition and subcellular distribution remains unknown. In particular, it is not known whether spinal neurons express functional kainate receptors and whether these receptors co-localize with AMPA receptors at postsynaptic sites. Rapid applications of kainate to cultured spinal cord neurons evoked desensitizing currents in the presence of GYKI 53655 (Fig. 3). Evidence that these currents are mediated by kainate receptors¹⁵ includes their blockade by SYM 2081 (Fig. 3a) and lanthanum (Fig. 3b), their insensitivity to cyclothiazide (Fig. 3c), and the reduction in their desensitization by exposure to concanavalin A (conA; Fig. 3d). Focal application of kainate (10 μ M) to spinal slices also elicited currents during continuous exposure to NMDA- and AMPA-receptor antagonists and conA ($n = 8$; 16.4 ± 2.7 pA; Fig. 3e). However, in the presence of CNQX (20 μ M) these currents were blocked ($n = 5$; 0.8 ± 1.0 pA; Fig. 3f). Thus, kainate receptors are available on postsynaptic cells to produce the kainate-receptor-mediated EPSCs.

In young rats, sensory neurons in the superficial dorsal horn of the spinal cord receive both low- and high-threshold afferents^{21,22}. To determine whether kainate and AMPA receptors co-exist at these

synapses, we used low-intensity stimulation, which activates a minimal number of presynaptic fibres. If kainate and AMPA receptors always co-localize, then EPSCs evoked by minimal stimulation should include both AMPA- and kainate-receptor-mediated components. However, only rapidly decaying EPSCs, typical of the AMPA-receptor-mediated component, were observed following low-intensity stimulation ($n = 7$; 8.7 ± 1.7 pA; $\tau = 3.5 \pm 0.5$ ms). Addition of GYKI 53655 (Fig. 4a; $n = 2$) or SYM 2206 ($n = 3$) blocked these EPSCs completely, even when a brief train of low-intensity stimulation was delivered^{3–5} (25 Hz, 6 pulses; $n = 5$; Fig. 4b). These results raise the possibility that kainate receptors might be displayed selectively at synapses formed by high-threshold afferents and not at synapses that receive low-threshold afferent inputs. Because accurate measurement of EPSC latency is difficult in thin-slice preparations, we studied the dependence of synaptic responses on stimulation intensity more systematically, using intensities that yield activation of different primary afferent fibre types in adult rats²³. The amplitude of SYM-2206-resistant EPSCs increased disproportionately with stimulus intensity (Fig. 4c–e). Thus, a significantly greater percentage of the EPSC was resistant to block by SYM 2206 during stimulation at high intensities, which were sufficient to activate A₈ and/or C fibres *in vivo*, than was the case for lower intensity stimulation (Fig. 4e).

At mossy-fibre inputs to hippocampal CA3 neurons, the demonstration of kainate-receptor-mediated EPSCs required stimulation with brief repetitive impulse trains^{3–5}. In contrast, at primary afferent synapses repetitive stimulation caused a frequency-dependent reduction in the amplitude of kainate-receptor-mediated EPSCs (0.05–5 Hz, data not shown; see Fig. 4f for 1-Hz stimulation). Pretreatment of primary afferent synapses with conA (5 μ M; Fig. 4f) abolished this depression, indicating that the depression may result from cumulative desensitization of postsynaptic kainate receptors.

To determine whether kainate receptors contribute to sensory transmission in ascending pathways, we injected a retrogradely transported fluorescent tracer into the thalamus and, two or three days later, recorded from labelled cells in spinal slices. Consistent with the results of a previous study²⁴, many cells in the superficial dorsal horn were labelled (Fig. 5a). Recordings were obtained from four labelled cells (for 10–15 min), all of which exhibited kainate-receptor-mediated EPSCs upon stimulation of the DREZ ($n = 3$) or dorsal root nerves ($n = 1$) (Fig. 5b).

Opiates produce antinociceptive effects by inhibiting sensory transmission in the dorsal horn⁶. Damgo (1 μ M), a selective μ -opioid-receptor agonist, blocked kainate-receptor-mediated EPSCs (Fig. 5c, d; $n = 5$). To study the role of kainate receptors in nociception further, we carried out behavioural experiments (Fig. 5e). Intrathecal administration of CNQX produced dose-dependent antinociception in the tail-flick and hot-plate tests. SYM 2206 had no effect in the hot-plate test but inhibited the tail-flick reflex at the highest dose. SYM 2081 was antinociceptive in both the tail-flick and the hot-plate tests. Latency in the cold-plate test²⁵ was not affected by any of the non-NMDA-receptor antagonists. The NMDA-receptor antagonists MK801 ($n = 6$; 7 nmol) and AP5 ($n = 4$; 25 nmol) increased latency in the cold-plate test, although motor side effects were also observed.

Our results show for the first time, to our knowledge, that spinal neurons express functional kainate receptors which contribute to synaptic transmission between primary afferent fibres and dorsal horn neurons. In contrast to previous results obtained with hippocampal mossy-fibre synapses, substantial kainate-receptor-mediated EPSCs were observed following individual, low-frequency, high-intensity stimuli. Low-threshold afferents appear to form synapses with few postsynaptic kainate receptors, whereas synapses that include both AMPA and kainate receptors are likely to receive inputs with a high activation threshold from, for example, C and/or A₈ fibres. Moreover, different subsets of postsynaptic glutamate

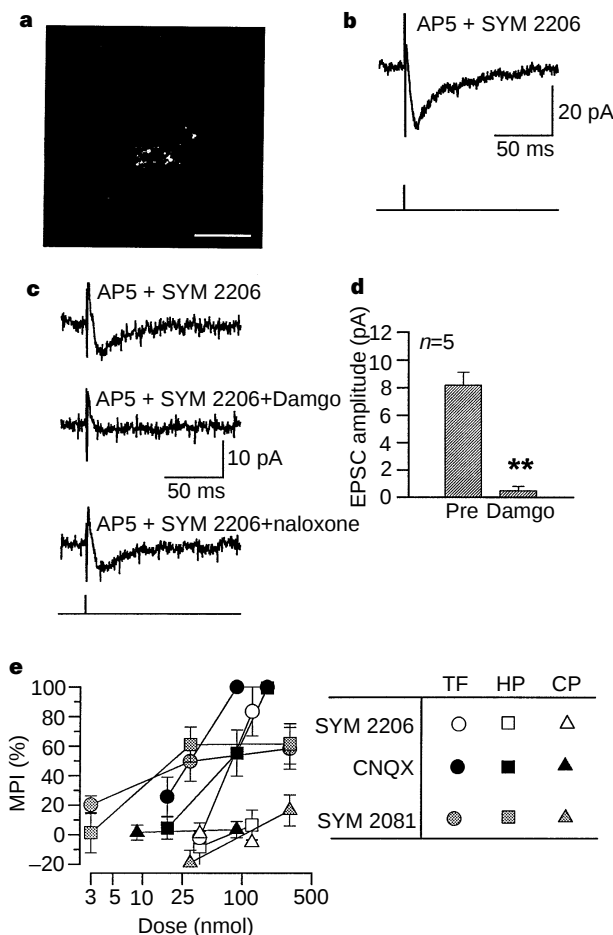


Figure 5 Kainate receptors contribute to ascending sensory transmission. **a**, A fluorescence photomicrograph of a dorsal horn cell that was labelled by fluorescent latex microspheres (light areas) 2 days after thalamic injection. Scale bar represents 20 μ m. **b**, An EPSC from a labelled spinothalamic tract cell in 100 μ M AP5 and 100 μ M SYM 2206. **c**, EPSCs in AP5 plus SYM 2206, in AP5, SYM 2206 and Damgo (1 μ M) and in AP5, SYM 2206 and naloxone (20 μ M). **d**, Mean amplitude ($\pm$ s.e.m.) of kainate EPSCs before and after application of Damgo in the presence of AP5 and SYM 2206 ($n = 5$). Double asterisk indicates $P < 0.001$. **e**, Antinociceptive effects of intrathecal administration of antagonists ($n = 3$ –7 for each dose). TF, tail flick; HP, hot plate; CP, cold plate.

matergic receptor may subserve behavioural responses to noxious heat and cold. Although glutamatergic synapses mediate pain sensation at all intensities, recent studies^{26,27} indicate that substance P and/or neurokinin A are important in sensing intense pain. Thus, it appears that both intensity and modality of pain are coded by different subtypes of postsynaptic receptor and by the release of different transmitters from primary afferent fibres. □

Methods

Spinal slices. Whole-cell recordings were made from spinal slices of rats at postnatal days 4–21 as described²⁸. EPSCs were evoked at 0.05 Hz with a bipolar tungsten electrode (stimulus width 0.1 or 0.4 ms) placed at the DREZ or dorsal root nerve. Monosynaptic EPSCs were identified as described²⁸. Only monosynaptic EPSCs were studied. Currents were filtered at 1 kHz and digitized at 5 kHz. Bicuculline methiodide (10 μ M) and strychnine hydrochloride (1 μ M) were added to the perfusion solution. Statistical comparisons were made using one-way analyses of variance (ANOVAs; Dunnett test for post-hoc comparison) or Student's *t*-test. *P* < 0.05 was considered to be significant. SYM 2206 and SYM 2081 were from Tocris–Cookson; other compounds were from RBI or Sigma.

Cultured neurons. Neurons were dissociated from the dorsal half of spinal cord slices and maintained for 7–14 days in culture using standard methods¹⁵. Whole-cell pipettes were filled with (in mM): 140 caesium glucuronate, 10 EGTA, 10 HEPES, 5 CsCl, 5 MgCl₂, 5 ATP and 1 GTP, pH 7.4. Drugs were dissolved in (in mM) 160 NaCl, 10 HEPES, 2 CaCl₂ plus 500 nM tetrodotoxin (pH 7.4), and applied by rapid local perfusion from a multibarrelled pipette as described¹⁵.

Labelling with fluorescent dye. Rats were anaesthetized with halothane (2–3%) delivered through a nose cone (with 30% O₂ balanced with N₂) and were positioned in a stereotaxic apparatus. Fluorescent tracer (0.5–1 μ l DiI, 0.2% in dimethylsulphoxide, or rhodamine latex microspheres) was injected into one side of the lateral and medial part of the thalamus²⁴. Two days after injection, the somata of ascending projection cells were visualized under epifluorescent illumination with a rhodamine filter set.

Behavioural tests. The tail-flick reflex and hot-plate (50 °C) and cold-plate (0 °C) tests were measured as described²⁵. Cold stimuli (0 °C) are believed to be noxious and to activate specific cold nociceptors²⁵. During intrathecal injection, mice or rats were anaesthetized with halothane (2%). Injections of drugs (mice, 5 μ l; rats, 15 μ l) were made with the steel tip of a 30-gauge needle connected by a 1-ft length of flexible PE-10 tubing to a 50- μ l syringe. Saline was used as a control. After the injection, it took 2–3 min for animals to recover. Data are presented as maximum possible inhibition (MPI) = (response latency – baseline response latency)/(cutoff time – baseline response latency) $\times$ 100.

Received 22 October; accepted 23 November 1998.

- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Lerma, J., Morales, M., Vicente, M. A. & Herreras, O. Glutamate receptors of the kainate type and synaptic transmission. *Trends Neurosci.* **20**, 9–12 (1997).
- Castillo, P. E., Malenka, R. C. & Nicoll, R. A. Kainate receptors mediate a slow synaptic current in hippocampal CA3 neurons. *Nature* **388**, 182–186 (1997).
- Vignes, M. & Collingridge, G. L. The synaptic activation of kainate receptors. *Nature* **388**, 179–182 (1997).
- Mulle, C. *et al.* Altered synaptic physiology and reduced susceptibility to kainate-induced seizures in GluR6-deficient mice. *Nature* **392**, 601–605 (1998).
- Yaksh, T. L. & Malmberg, A. B. In *Textbook of Pain* (edited by Wall, P. D. & Melzack, R.) 165–200 (Churchill Livingstone, New York, 1994).
- Yoshimura, M. & Jessell, T. M. Amino acid-mediated EPSPs at primary afferent synapses with substantia gelatinosa neurones in the rat spinal cord. *J. Physiol. (Lond.)* **430**, 315–335 (1990).
- Kumazawa, T. & Perl, E. R. Excitation of marginal and substantia gelatinosa neurons in the primate spinal cord: indications of their place in dorsal horn function organization. *J. Comp. Neurol.* **177**, 417–434 (1978).
- Light, A. R., Trevino, D. L. & Perl, E. R. Morphological features of functionally defined neurons in the marginal zone and substantia gelatinosa of the spinal dorsal horn. *J. Comp. Neurol.* **186**, 151–171 (1979).
- Bleakman, D. *et al.* Activity of 2,3-benzodiazepines at naive rat and recombinant human glutamate receptors *in vitro*: stereospecificity and selectivity profiles. *Neuropharmacol.* **35**, 1689–1702 (1996).
- Pelletier, J. C., Hesson, D. P., Jones, K. A. & Costa, A.-M. Substituted 1,2-dihydrophthalazines: potent, selective, and noncompetitive inhibitors of the AMPA receptor. *J. Med. Chem.* **39**, 343–346 (1996).
- Patrain, A. V., Morales, M. & Lerma, J. Selective antagonism of AMPA receptors unmasks kainate receptor-mediated responses in hippocampal neurons. *Neuron* **14**, 185–189 (1995).
- Wilding, T. J. & Huettner, J. E. Differential antagonism of alpha-amino-5-hydroxy-5-methyl-4-isoxazolepropionic acid-preferring and kainate-preferring receptors by 2,3-benzodiazepines. *Mol. Pharmacol.* **47**, 582–587 (1995).

- Jones, K. A., Wilding, T. J., Huettner, J. E. & Costa, A.-M. Desensitization of kainate receptors by kainate, glutamate and diastereomers of 4-methylglutamate. *Neuropharmacol.* **36**, 853–863 (1997).
- Wilding, T. J. & Huettner, J. E. Activation and desensitization of hippocampal kainate receptors. *J. Neurosci.* **17**, 2713–2721 (1997).
- Partin, K. M., Patneau, D. K., Winters, C. A., Mayer, M. L. & Buonanno, A. Selective modulation of desensitization at AMPA versus kainate receptors by cyclothiazide and concanavalin A. *Neuron* **11**, 1069–1082 (1993).
- Todd, A. J., Spike, R. C., Price, R. F. & Neilson, M. Immunocytochemical evidence that neurotension is present in glutamatergic neurons in the superficial dorsal horn of the rat. *J. Neurosci.* **14**, 774–784 (1994).
- Tölle, T. R., Berthele, A., Zieglgänsberger, W., Seeburg, P. H. & Wisden, W. The differential expression of 16 NMDA and non-NMDA receptor subunits in the rat spinal cord and in periaqueductal gray. *J. Neurosci.* **13**, 5009–5028 (1993).
- Wisden, W. & Seeburg, P. H. A complex mosaic of high-affinity kainate receptors in rat brain. *J. Neurosci.* **13**, 3582–3598 (1993).
- Petralia, R. S., Wang, Y.-X. & Wenthold, R. J. Histological and ultrastructural localization of the kainate receptor subunits, KA2 and GluR6/7, in the rat nervous system using selective antipeptide antibodies. *J. Comp. Neurol.* **349**, 85–110 (1994).
- Fitzgerald, M., Butcher, T. & Shortland, P. Development changes in the laminar termination of A fibre cutaneous sensory afferents in the rat spinal cord dorsal horn. *J. Comp. Neurol.* **348**, 225–233 (1994).
- Coggeshall, R. E., Jennings, E. A. & Fitzgerald, M. Evidence that large myelinated primary afferent fibers make synaptic contacts in lamina II of neonatal rats. *Brain Res. Dev. Brain Res.* **92**, 81–90 (1996).
- Baba, H., Yoshimura, M., Nishi, S. & Shimoji, K. Synaptic responses of substantia gelatinosa neurons to dorsal column stimulation in rat spinal cord *in vitro*. *J. Physiol. (Lond.)* **478**, 87–99 (1994).
- Huang, L.-Y. M., Carlton, S. M. & Willis, W. D. Identification of spinothalamic tract cells in fresh, unfixed rat spinal cord. *J. Neurosci. Methods* **14**, 91–96 (1985).
- Zhuo, M. NMDA receptor-dependent long term hyperalgesia after tail amputation in mice. *Eur. J. Pharmacol.* **349**, 211–220 (1998).
- Cao, Y. Q. *et al.* Primary afferent tachykinins are required to experience moderate to intense pain. *Nature* **392**, 390–394 (1998).
- De Felipe, C. *et al.* Altered nociception, analgesia and aggression in mice lacking the receptor for substance P. *Nature* **392**, 394–397 (1998).
- Li, P. & Zhuo, M. Silent glutamatergic synapses and nociception in mammalian spinal cord. *Nature* **393**, 695–698 (1998).

Acknowledgements. We thank D. Leander for GYKI 53655. This work was supported in part by grants from NIDA (to M.Z.) and NINDS (to J.E.H.) of the NIH.

Correspondence and requests for materials should be addressed to M.Z. (e-mail: zhuom@morpheus.wustl.edu).

The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus

Jacqueline J. L. Jacobs*, Karin Kieboom*, Silvia Marino†, Ronald A DePinho‡ & Maarten van Lohuizen*

* Division of Molecular Carcinogenesis and † Division of Molecular Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

‡ Dana Farber Cancer Institute, Harvard Medical School, 44 Binney Street (M463), Boston, Massachusetts 02115, USA

The *bmi-1* gene was first isolated as an oncogene that cooperates with *c-myc* in the generation of mouse lymphomas^{1,2}. We subsequently identified *Bmi-1* as a transcriptional repressor belonging to the mouse Polycomb group^{3–6}. The Polycomb group comprises an important, conserved set of proteins that are required to maintain stable repression of specific target genes, such as homeobox-cluster genes, during development^{7–9}. In mice, the absence of *bmi-1* expression results in neurological defects and severe proliferative defects in lymphoid cells, whereas *bmi-1* overexpression induces lymphomas^{4,10}. Here we show that *bmi-1*-deficient primary mouse embryonic fibroblasts are impaired in progression into the S phase of the cell cycle and undergo premature senescence. In these fibroblasts and in *bmi-1*-deficient lymphocytes, the expression of the tumour suppressors p16 and p19^{Arf}, which are encoded by *ink4a*, is raised markedly. Conversely, overexpression of *bmi-1* allows fibroblast immortalization, downregulates expression of p16 and p19^{Arf} and, in combination with *H-ras*, leads to neoplastic transformation. Removal of *ink4a* dramatically reduces the lymphoid and neurological defects seen in *bmi-1*-

deficient mice, indicating that *ink4a* is a critical *in vivo* target for Bmi-1. Our results connect transcriptional repression by Polycomb-group proteins with cell-cycle control and senescence.

To investigate the role of *bmi-1* in cell proliferation in biochemical detail, we used primary mouse embryonic fibroblasts (MEFs) derived from *bmi-1*^{-/-} embryos. *Bmi-1*^{-/-} MEFs have a significantly reduced proliferation rate compared with *bmi-1*^{+/+} MEFs (Fig. 1a). This is because of an S-phase defect, as shown by a reduced rate of incorporation of the DNA-labelling dye bromodeoxyuridine (BrdU) (Fig. 1b). Cytoplasmic enlargement, flattening of the cells, unresponsiveness to growth factors and expression of acidic β -galactosidase¹¹ in 80% of arrested passage 3 *bmi-1*^{-/-} MEFs indicates premature entry into senescence, in contrast to wild-type cells which first arrest at passage 7 (Fig. 1e). Re-expression of wild-type *bmi-1* by retroviral transduction, but not expression of *bmi-1* with a mutation in the RING-finger domain that is incapable of inducing tumours *in vivo*¹⁰, prevents premature senescence and completely restores the proliferative capacity of *bmi-1*^{-/-} cells, indicating specificity of *bmi-1*^{-/-} arrest (Fig. 1c, d). Wild-type MEFs that overexpress *bmi-1* proliferate significantly faster and to higher cell densities than control cells, and have an extended lifespan: they either become immortal immediately or enter a slow growth period at about passage 13, after which they easily become immortalized (Fig. 1c and Supplementary Information). Similar data were obtained using a defined 3T3 culture scheme (see Supplementary Information). Thus, lack of *bmi-1* expression causes premature entry into senescence, whereas overexpression of *bmi-1* allows immortalization.

We observed a severe downregulation of cyclin A and cyclin E

protein levels in *bmi-1*^{-/-} MEFs, which did not occur in *bmi-1*^{-/-} MEFs infected with the *bmi-1* retrovirus (Fig. 2a). Although no differences were observed in amounts of the cyclin-dependent kinase inhibitors p21 and p27 or the tumour suppressor p53, an upregulation of p16 protein in *bmi-1*^{-/-} MEFs compared with wild-type MEFs was clearly seen (Fig. 2a, b, and data not shown). Conversely, overexpression of *bmi-1*, but not of the RING-finger mutant protein, led to rapid downregulation of p16 levels in *bmi-1*^{-/-} MEFs and in wild-type MEFs (Fig. 2a, b). Significantly, p16 is thought to be involved in the induction of replicative senescence¹²⁻¹⁴. In agreement with the highly conserved structure and function of Polycomb-group (PcG) genes^{7,8}, overexpression of *bmi-1* also downregulates p16 protein levels in TiG-3 human primary fibroblasts (Fig. 2c). This downregulation is accompanied by delayed entry into senescence, although the *bmi-1*-overexpressing human cells are not fully immortalized and arrest after ± 7 more population doublings than normal (Fig. 2d). Human cells show more complex and multifactorial regulation of the senescence checkpoint than murine cells^{12,13,15,16}. In agreement with this we observed no detectable telomerase reactivation upon *bmi-1* overexpression in TiG-3 cells using the sensitive telomeric-repeat-amplification protocol (data not shown).

The tumour suppressor p19^{Arf}, which, like p16, is also encoded by the *ink4a* locus but is controlled by a separate promoter, is also important in cell proliferation and senescence¹⁷. p16 and, to a lesser extent, p19^{Arf} steady-state transcript levels were upregulated on average eightfold and two- to threefold, respectively, in *bmi-1*^{-/-} MEFs (Fig. 3a) and splenocytes (Fig. 4a), whereas the neighbouring

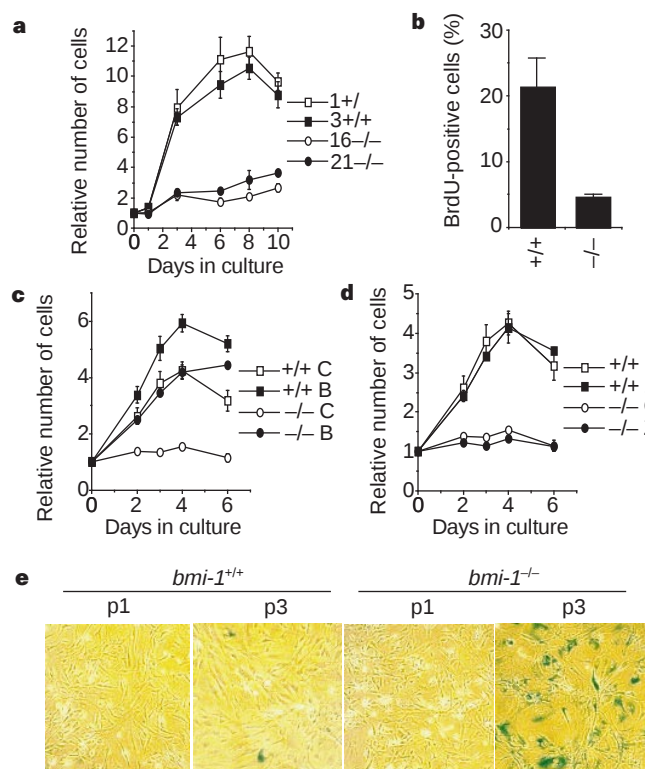


Figure 1 Impaired proliferation of *bmi-1*^{-/-} MEFs is rescued by wild-type but not mutant *bmi-1*. **a**, Growth curves for two independent *bmi-1*^{+/+} and *bmi-1*^{-/-} MEF cultures at passage 2. The numbers refer to independently derived MEF cultures. **b**, Level of BrdU incorporation by asynchronously growing passage 3 *bmi-1*^{+/+} and *bmi-1*^{-/-} MEFs. **c, d**, Growth curves for passage 2 *bmi-1*^{+/+} and *bmi-1*^{-/-} MEFs infected at passage 1 with 'empty' control LZRS retrovirus (C), LZRS-*bmi-1* virus (B) or LZRS-BZN virus (Z): wild type Bmi-1 rescues proliferation, whereas the BZN Bmi-1 RING-finger mutant is unable to do so. **e**, β -Galactosidase (pH 6) activity in passage 1 and 3 *bmi-1*^{+/+} and *bmi-1*^{-/-} MEFs.

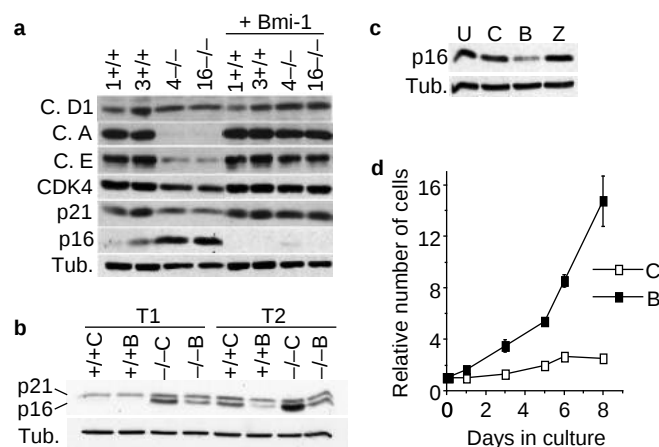


Figure 2 Effects of absence or overexpression of *bmi-1* on cell-cycle-regulatory proteins in primary fibroblasts. **a-c**, Western blots of cell lysates. **a**, Cell lysates from control- or LZRS-*bmi-1*-infected passage 3 *bmi-1*^{+/+} and *bmi-1*^{-/-} MEFs. Tubulin (Tub.) serves as loading control. C.D1, cyclin D1; C.A, cyclin A; C.E, cyclin E; CDK4, cyclin-dependent kinase 4. **b**, Cell lysates from control (C) or LZRS-*bmi-1* (B)-infected *bmi-1*^{+/+} and *bmi-1*^{-/-} MEFs, showing p16 and p21 expression 48 h (T1) and 6 d (T2) after infection. **c**, Downregulation of p16 protein in primary human TiG-3 cells after infection with LZRS-*bmi-1* virus (B), but not after infection with control (C) or BZN-mutant (Z) virus. U, uninfected. **d**, Growth curves for TiG-3 cells infected at PDL 50 with control (C) or Bmi-1-expressing (B) retrovirus; growth was monitored for control-infected cells starting at PDL 68 (nearly senescent) and for Bmi-1-overexpressing cells at PDL 70 (still proliferating).

p15^{ink4b} gene was only modestly affected (1.6-fold upregulation; Fig. 3a). This gradual effect on the *ink4a* locus is in line with transcriptional repression by PcG proteins in *Drosophila*, which is thought to spread over a limited distance from *cis* elements into adjacent chromatin^{18–20}. As PcG proteins act in multiprotein complexes^{5,21,22}, we analysed senescence entry by MEFs deficient in another PcG gene. Significant upregulation of p16 messenger RNA levels (threefold) was also observed in *mel18*^{-/-} MEFs²³; this upregulation was accompanied by premature senescence entry at passage 5 (Fig. 3b). The passage of arrest correlated well with relative levels of p16 induction. This indicates that the effects on the cell cycle and senescence are intrinsic to the interaction of PcG proteins in repressive complexes. Consistent with transcriptional repression by PcG proteins, downregulation of steady-state levels of p16 (fourfold) and p19^{Arf} (2.4-fold) mRNA were easily detected in *bmi-1*-infected MEFs, but not in cells infected with control virus (Fig. 4b).

Unlike wild-type MEFs but as reported for *ink4a*^{-/-} MEFs, MEFs overexpressing *bmi-1* could be transformed by oncogenic *ras* alone, as detected by a focus-formation assay; serial dilutions of the *bmi-1* retrovirus led to a dose-dependent reduction in the number of transformed foci (Fig. 4c). This indicates that *bmi-1* and *ras* cooperate in neoplastic transformation of primary MEFs; in a similar way, *ras* and the loss of p16 and p19^{Arf} expression cooperate to induce neoplastic transformation^{17,24}.

To determine *in vivo* to what extent the upregulation of p16 and p19^{Arf} is needed to produce the observed proliferative defects in *bmi-1*^{-/-} mice, mice heterozygous for the *bmi-1* gene and for exons 2 and 3 of *ink4a* were intercrossed to generate double-knockout mice as well as heterozygous and wild-type control littermates. In the first litters (37 offspring out of 5 litters), we observed two mice of an intermediate size, two very small mice and a large group of mice of normal size. The two small mice developed ataxia at about 3–4 weeks after birth, unlike the intermediate-sized mice. Genotyping confirmed that the two small mice were *bmi-1*^{-/-} *ink4a*^{+/-} whereas the mice of intermediate size were *bmi-1*^{-/-} *ink4a*^{-/-}. At 5 weeks of age, the *bmi-1*^{-/-} *ink4a*^{+/-} mice had progressed to severe ataxia, whereas the two double-knockout mice were still indistinguishable in behaviour from wild-type mice.

Histopathological analysis of *bmi-1*^{-/-} *ink4a*^{-/-} mice showed a dramatic rescue of the cerebellar defects normally observed in *bmi-1*^{-/-} mice⁴ (Fig. 5a). The size of the cerebellum was comparable to the wild-type cerebellum, the width of the granular cell layer was restored and the cellularity in the granular and molecular layers increased (Fig. 5b). Notwithstanding the rescued cellularity, the reactive astrogliosis observed in *bmi-1*^{-/-} mice⁴ was only slightly reduced in the double-knockout mice (see Supplementary Information). This result indicates that there may be only a partial rescue of the neurological defects, resulting in a delayed onset and slower progression of the pathological process. These results, which are

in line with the significant upregulation of p16 and p19^{Arf} expression in RNA isolated from *bmi-1*^{-/-} brains (data not shown) and with the absence of behavioural disorders in double-knockout mice at 5 weeks of age, indicate that *ink4a* is a critical target for *bmi-1* in this pathological setting too. Significant rescue of thymocyte and splenocyte cell numbers to up to 50–70% of wild-type numbers was seen in *bmi-1*^{-/-} *ink4a*^{-/-} mice, in contrast with *bmi-1*^{-/-} *ink4a*^{+/-} littermates which retained only 2–4% of wild-type cell numbers (Fig. 5c). Fluorescence-activated cell sorting (FACS) analysis, using standard T- and B-cell differentiation cell-surface markers, of *bmi-1*^{-/-} *ink4a*^{-/-} thymocytes and splenocytes showed profiles indistinguishable from those of wild-type cells; *bmi-1*^{-/-} *ink4a*^{+/-} littermates, however, showed increased populations of CD4⁺ CD8⁻, CD3⁺ CD25⁺ or sIg⁺ B220⁺ immature cells (Fig. 5d). Independent crosses confirmed the rescue of cerebellar and lymphoid defects in another *ink4a*^{-/-} *bmi-1*^{-/-} mouse. Full rescue of the proliferative defects and premature senescence entry of *bmi-1*^{-/-} MEFs was also observed in *bmi-1*^{-/-} *ink4a*^{-/-} MEFs, which grew

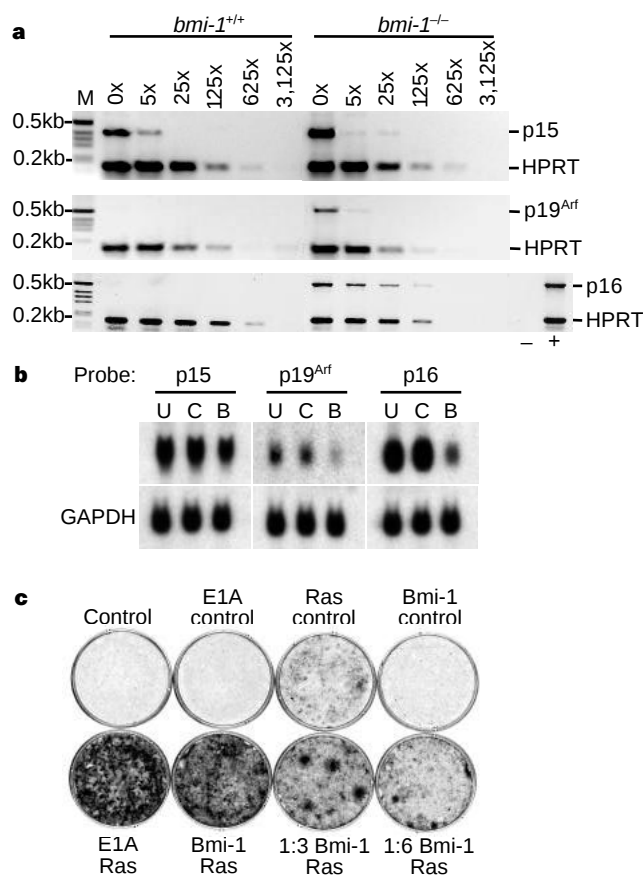


Figure 4 p15, p16 and p19^{Arf} mRNA levels in *bmi-1*^{+/+} and *bmi-1*^{-/-} splenocytes and control or *bmi-1*-virus-infected MEFs, and cooperation between Ras and Bmi-1 in neoplastic transformation. **a**, Semiquantitative RT-PCR analysis of specific p15, p16 and p19^{Arf} transcripts in *bmi-1*^{+/+} and *bmi-1*^{-/-} spleens. Simultaneous amplification of hypoxanthine guanine phosphoribosyl transferase (Hprt) served as an internal control. Water (–) and total MEF cDNA (+) also served as controls. M, markers; kb, kilobases. The numbers above the lanes represent serial dilution of first-strand cDNA template used. **b**, Northern blot analysis of p15, p19^{Arf} exon 1β and p16 exon 1α transcripts in uninfected (U), control (C) or LZRS-*bmi-1*-infected (B) *bmi-1*^{-/-} MEFs. GAPDH served as a loading control. **c**, Focus-formation assay using *bmi-1*^{+/+} MEFs infected with 'empty' retrovirus (top left) or with retroviruses encoding the proteins indicated. The plates in the top row, except top left, were infected with retroviruses encoding 12S E1A, Ras or Bmi-1, and were superinfected with an 'empty' retrovirus. The plates in the bottom row were infected with retroviruses encoding 12S E1A, undiluted Bmi-1, or Bmi-1 diluted 1:3 or 1:6; they were then superinfected with H-ras V12 retrovirus.

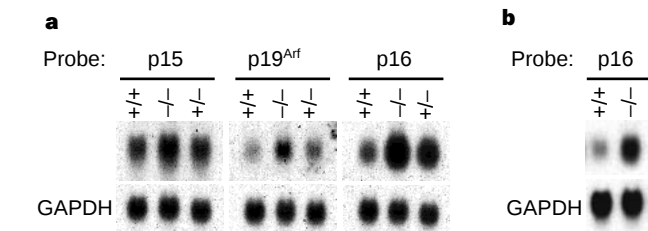


Figure 3 p15, p16 and p19^{Arf} mRNA levels in *bmi-1*^{+/+}, *bmi-1*^{-/-} and *bmi-1*^{+/-} MEFs and in *mel-18*^{+/+} and *mel-18*^{-/-} MEFs. **a**, Northern blot analysis of p15, p19^{Arf} exon 1β and p16 exon 1α transcripts in passage 3 *bmi-1*^{+/+}, *bmi-1*^{-/-} or *bmi-1*^{+/-} MEFs. **b**, p16 mRNA levels in passage 3 *mel-18*^{+/+} and *mel-18*^{-/-} MEFs. GAPDH was included as a loading control. All signals were quantified relative to the loading control using a phosphorimager.

faster, readily bypassed senescence arrest and were indistinguishable from *ink4a*^{-/-} MEFs²⁴ in this respect (Fig. 5e).

The overall improved health status of the double-knockout animals, the normal lymphocyte FACS profiles and the increased cellularity of the cerebellar layers indicate a remarkable restoration of lymphoid homeostasis and neurological functions. This, together with the results of our studies of MEFs, provides evidence that a

common, critical cell-cycle-control locus (*ink4a*) lies at the heart of proliferative defects in very different affected cell types in *bmi-1*^{-/-} mice. As expected for a major downstream target, overexpression of either p16 or p19^{Arf} from a retroviral vector in MEFs already overexpressing *bmi-1* still causes cell-cycle arrest (data not shown), indicating that *bmi-1* overexpression does not lead to major activations of signalling pathways downstream of *ink4a*. In good agreement with *ink4a*-mediated arrest acting before M₂ crisis, karyotype analysis of arrested *bmi-1*^{-/-} MEFs showed no obvious abnormalities and no significant reduction in telomere length was observed by fluorescence *in situ* hybridization analysis (data not shown). These results, together with our observations of human primary cells, indicate that the wearing away of telomeres or telomerase activity are not involved in the observed *ink4a*-mediated arrest in *bmi-1*^{-/-} cells.

Our results place *bmi-1* at an early step in tumorigenesis as an immortalizing oncogene that is capable of cooperating with *H-ras* in transformation of primary cells, and provide a convenient *in vitro* assay for the effects of PcG dose on cell-cycle regulation. Moreover, our results offer a plausible explanation for the strong cooperation between *bmi-1* and *c-myc* or *H-ras*. Whereas oncogenes such as *H-ras*, *c-myc* and *E1A* induce p19^{Arf}/p53 (refs 25, 26) and/or p16 (ref. 15) in primary cells to prevent immortalization during ectopic mitogenic signalling, *bmi-1* acts primarily by suppressing p16/cyclin D/retinoblastoma protein (Rb) and/or p19^{Arf}/MDM2/p53 tumour-suppressor pathways, thereby allowing progression through the cell cycle. Both of these latter tumour-suppressor pathways, when defective, cooperate efficiently with *c-myc* or activated cyclin E in the induction of mouse lymphomas^{27,28}. The efficient *in vivo* cooperation of these oncogenes, and our observations that proliferative defects in *c-myc*^{-/-} rat TGR-1 cells²⁹ cannot be rescued by overexpression of *bmi-1* and that *bmi-1*^{-/-} MEFs cannot be rescued by *c-myc*-expressing retroviruses, indicates that these oncogenes may fulfil separate, cooperative functions in immortalization (data not shown).

In conclusion, our *in vivo* and *in vitro* results identify the tumour suppressors p16 and p19^{Arf} as critical downstream targets for the PcG gene and oncogene *bmi-1*, with regard to its effects on cell proliferation and senescence. This establishes for the first time, to our knowledge, a connection between PcG-mediated silencing, cell-cycle regulation and the senescence checkpoint, and shows that PcG-mediated transcription repression is not only crucial for regulation of *Hox* genes, but also controls the expression of critical cell-cycle regulators. PcG-dependent regulation of *ink4a* may reflect the fact that two of the most frequently disrupted tumour-suppressor pathways³⁰ in a variety of neoplasias, p16/cyclin D/Rb and p19^{Arf}/MDM2/p53, each include a protein encoded by the same gene, *ink4a*¹⁶. Thus a tight and coordinated transcriptional regulation of this locus is required. The significance of regulation of *ink4a* by *bmi-1* is underscored by our observation that this regulation is conserved in primary MEFs and human fibroblasts, indicating that overexpression of Bmi-1 may contribute to human neoplasias that retain wild-type *ink4a*. □

Methods

Cell culture, growth curves and retroviral infection. Cells were maintained in DMEM medium supplemented with 10% fetal bovine serum (Gibco). Head and organs of day 14.5 embryos were dissected; fetal tissue was rinsed in PBS, minced, rinsed twice in PBS and kept overnight on ice with 100 µl trypsin/EDTA (Gibco). The next day another 100 µl trypsin/EDTA was added and fetal tissue was incubated for 30 min at 37 °C and subsequently dissociated in medium. After removal of large tissue clumps, the remaining cells were plated out in a 175 cm² flask. After 48 h, confluent cultures were frozen down. These cells were considered as being passage 1 MEFs. For continuous culturing, MEF cultures were split 1:3, 1 passage being equivalent to 1.8 population doublings (PDLs). TiG-3 primary human diploid fibroblast cultures were split 1:4, 1 passage being comparable to 2 PDLs. For growth curves, 2.5 × 10⁴ cells were

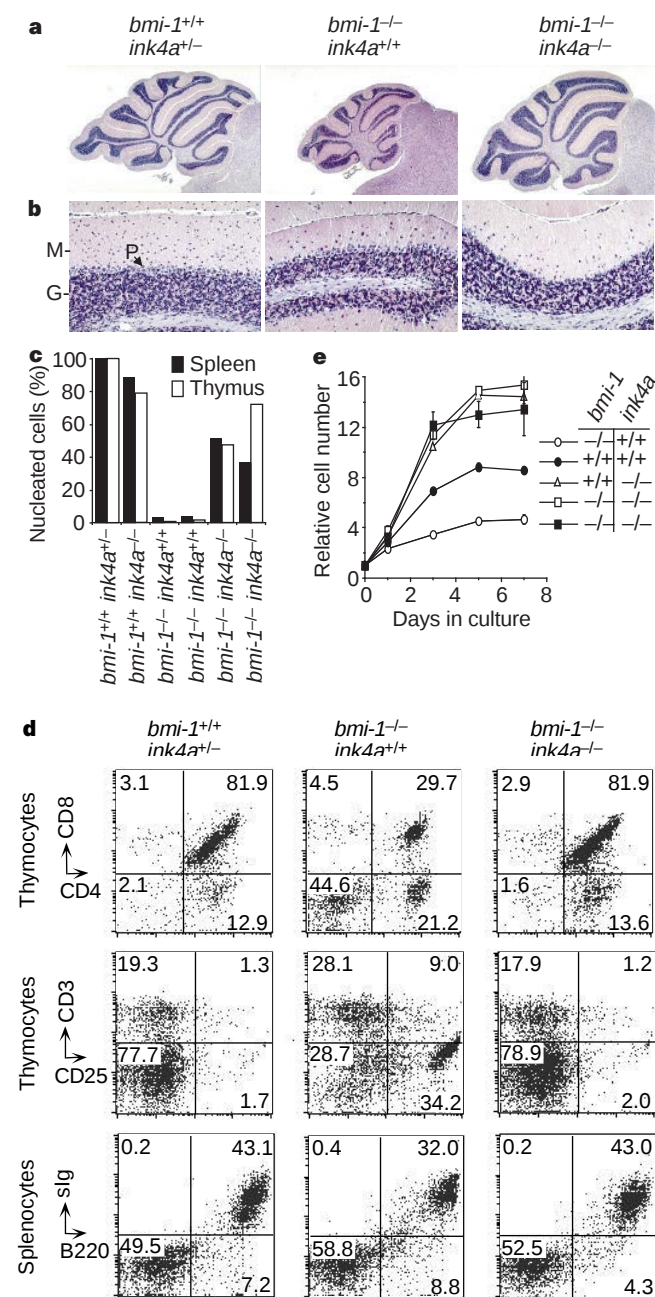


Figure 5 Rescue of *bmi-1*^{-/-}-associated proliferative defects in *bmi-1*^{-/-} *ink4a*^{-/-} mice and MEFs. **a, b**, Haematoxylin-stained cerebellar sagittal sections. **a**, Cerebellum of 5-week-old *bmi-1*^{+/+} *ink4a*^{+/+}, *bmi-1*^{-/-} *ink4a*^{+/+} and *bmi-1*^{-/-} *ink4a*^{-/-} littermates, showing up to 90% restoration of cellularity in double knockouts. **b**, Higher magnification of molecular (M) and granular (G) layers, showing increased cellularity and more regular spacing of Purkinje cells (P) in *bmi-1*^{-/-} *ink4a*^{-/-} mice compared with *bmi-1*^{-/-} *ink4a*^{+/+} mice. **c**, Per cent nucleated cells in spleen and thymus of mice generated by intercrossing *bmi-1*^{+/+} *ink4a*^{-/-} mice. **d**, Flow-cytometric analysis of thymocytes and splenocytes of 'wild-type' *bmi-1*^{+/+} *ink4a*^{+/+}, *bmi-1*^{-/-} *ink4a*^{+/+} and *bmi-1*^{-/-} *ink4a*^{-/-} mice. Note the almost complete restoration to wild-type levels of B- and T-lymphocyte subpopulations in double knockouts. **e**, Growth curves: *bmi-1*^{-/-} *ink4a*^{-/-} MEFs proliferate like *bmi-1*^{+/+} *ink4a*^{-/-} MEFs.

plated in triplicate into 12-well plates. At various time points cells were stained with crystal violet (Sigma) and the optical density at 590 nm was determined¹⁵. Values were normalized to the optical density at day 0 (20 h after plating). Phoenix producer cells were used to generate retroviral stocks as described¹⁵. For infection, subconfluent passage 1 MEF cultures were incubated at 37 °C with viral supernatant in the presence of 4 µg ml⁻¹ polybrene (Sigma). After 6 h the viral supernatant was diluted 1:3 with complete medium and left on the cells for 42 h. Appropriate dilutions of viral supernatants were used such that 100% of MEFs were infected.

BrdU-incorporation assay and western blotting. MEFs were grown on coverslips and incubated for 4 h with 10 µM BrdU (Amersham). Cells were washed in PBS, fixed for 15 min at -20 °C in 5% acetic acid/95% ethanol and incubated in PBS. Fixed cells were incubated for 20 min in 2 M HCl/0.5% Triton-X100, for 30 min in blocking solution (5% fetal calf serum, 5% normal goat serum in PBS/0.02% Triton-X100), for 1 h with 1:10 diluted anti-BrdU monoclonal antibody (DAKO) in blocking solution, and overnight at 4 °C with 1:50 diluted fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse antibody (Jackson Immuno Research Labs) and 4,6-diamidino-2-phenylindole (DAPI) in blocking solution; this was followed by 3 washes for 5 min each in PBS/0.02% Triton-X100. Cells were embedded in Vectastain (Vector Labs) and the percentage of BrdU-labelled cells (FITC:DAPI ratio) was quantified using a fluorescence microscope. For protein analysis, cells were washed with PBS, scraped and lysed on ice in RIPA buffer (150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% DOC, 50 mM Tris-HCl, pH 8.0, 2 mM EDTA, pH 8.0, 0.2 mM phenylmethylsulphonylfluoride (PMSF), 0.5 mM dithiothreitol (DTT)). Cleared lysates were assayed for protein concentration. Equal amounts of protein were separated on 12.5% SDS-PAGE and transferred to nitrocellulose. Western blot analysis was according to standard methods using enhanced chemiluminescence (Amersham). A list of antisera used is available on request.

Expression analysis. Total RNA was extracted using guanidinium thiocyanate, separated on 1.2% agarose, transferred to nitrocellulose and hybridized according to standard procedures with ³²P-labelled probes specific for exon 1α of mouse p16, for exon 1β of mouse p19^{Arf}, for mouse p15 or for rat glyceraldehyde-3-phosphate dehydrogenase (GAPDH). For semiquantitative reverse transcription with polymerase chain reaction (RT-PCR), first-strand complementary DNA was generated from 1 µg total RNA using Superscript II RT (Gibco) and oligo-dT primer according to the manufacturer's instructions. Primer sequences are available upon request. PCR reactions were performed on five-times serial dilutions of first-strand cDNA in 50 µl containing 1× Taq PCR buffer, 1.5 mM MgCl₂, 200 µM dNTPs, 0.5 µM of each of four primers, 1 µl of first-strand cDNA template and 1.25 units of Taq DNA Polymerase (Gibco), using 35 cycles of denaturation (94 °C, 1 min), annealing (60 °C, 45 s) and extension (72 °C, 2 min). Products were resolved on 2% NuSieve agarose gels.

Generation of *bmi-1*^{-/-} *ink4a*^{-/-} mice. *bmi-1*^{+/-} FVB mice⁴ and *ink4a*^{+/-} mice on a mixed 129/Sv; C57BL/6 genetic background²⁴ were crossed to generate *bmi-1*^{+/-} *ink4a*^{+/-} mice, which were subsequently intercrossed to generate double-knockout offspring together with control littermates. Mice were genotyped routinely by PCR or Southern blot analysis of tail DNA.

Cell count and flow-cytometric analysis. Cell suspensions of lymphoid organs were prepared by mincing the tissue though an open filter chamber. Cell suspensions were depleted of erythrocytes and the number of nucleated cells was determined with a Casy-1 TT automated cell counter (Schäfer, Reutlingen, Germany). Flow cytometry with standard B- and T-cell differentiation was done nearly as described¹⁰.

Histological analysis. Brains were fixed in 4% formaldehyde in PBS, paraffin-embedded and cut into 4-µm serial sagittal sections. Sections at different levels were stained with haematoxylin and eosin.

Received 10 September; accepted 12 November 1998.

1. van Lohuizen, M. *et al.* Identification of cooperating oncogenes in Eµ-*myc* transgenic mice by provirus tagging. *Cell* **65**, 735–752 (1991).
2. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in Eµ-*myc* transgenic mice. *Cell* **65**, 753–763 (1991).
3. van Lohuizen, M., Frasch, M., Wientjens, E. & Berns, A. Sequence similarity between the mammalian *Bmi-1* proto-oncogene and the *Drosophila* regulatory genes *Psc* and *Su(z)2*. *Nature* **353**, 353–355 (1991).
4. van der Lugt, N. M. T. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *Bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–769 (1994).
5. Alkema, M. J. *et al.* *MPc2*, a new murine homologue of the *Drosophila* Polycomb protein is a member of the mouse Polycomb transcriptional repressor complex. *J. Mol. Biol.* **273**, 993–1003 (1997).

6. Alkema, M. J., van der Lugt, N. M. T., Bobeldijk, R. C., Berns, A. & van Lohuizen, M. Transformation of axial skeleton due to overexpression of *Bmi-1* in transgenic mice. *Nature* **374**, 724–727 (1995).
7. Paro, R. Propagating memory of transcriptional states. *Trends Genet.* **11**, 295–298 (1995).
8. van Lohuizen, M. Functional analysis of mouse Polycomb-group genes. *Cell. Mol. Life Sci.* **54**, 71–79 (1998).
9. Gould, A. Functions of mammalian Polycomb group and trithorax group related genes. *Curr. Opin. Genet. Dev.* **7**, 488–494 (1997).
10. Alkema, M. J., Jacobs, H., van Lohuizen, M. & Berns, A. Perturbation of B and T cell development and predisposition to lymphomagenesis in Eµ-*Bmi1* transgenic mice require the *Bmi1* RING finger. *Oncogene* **15**, 899–910 (1997).
11. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
12. Hara, E. *et al.* Regulation of p16 CDKN2 expression and its implications for cell immortalisation and senescence. *Mol. Cell. Biol.* **16**, 859–867 (1996).
13. Alcorta, D. A. *et al.* Involvement of the cyclin-dependent kinase inhibitor p16 (INK4a) in replicative senescence of normal human fibroblasts. *Proc. Natl Acad. Sci. USA* **93**, 13742–13747 (1996).
14. Zindy, F., Quelle, D. E., Roussel, M. F. & Sherr, C. J. Expression of the p16ink4a tumour suppressor versus other ink4 family members during development and aging. *Oncogene* **15**, 203–211 (1997).
15. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16ink4a. *Cell* **88**, 593–602 (1997).
16. Haber, D. A. Splicing into senescence: the curious case of p16 and p19^{Arf}. *Cell* **91**, 555–558 (1997).
17. Kamijo, T. *et al.* Tumour suppression at the mouse *ink4a* locus mediated by the alternative reading frame product p19^{Arf}. *Cell* **91**, 649–659 (1997).
18. Strutt, H., Cavalli, G. & Paro, R. Colocalisation of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* **12**, 3621–3632 (1997).
19. Cavalli, G. & Paro, R. The *Drosophila* *Fab-7* chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505–518 (1998).
20. Pirota, V. PcG complexes and chromatin silencing. *Curr. Opin. Genet. Dev.* **7**, 249–258 (1997).
21. Alkema, M. J. *et al.* Identification of *Bmi1*-interacting proteins as constituents of a multimeric mammalian Polycomb complex. *Genes Dev.* **11**, 226–240 (1997).
22. Gunster, M. J. *et al.* Identification and characterisation of interactions between the vertebrate Polycomb-group protein BMI1 and the human homologues of Polyhomeotic. *Mol. Cell. Biol.* **17**, 2326–2335 (1997).
23. Akasaka, T. *et al.* A role for *mel-18*, a *Polycomb* group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* **122**, 1513–1522 (1996).
24. Serrano, M. *et al.* Role of the *ink4a* locus in tumour suppression and cell mortality. *Cell* **85**, 27–37 (1996).
25. Zindy, F. *et al.* Myc signalling via the ARF tumour suppressor regulates p53-dependent apoptosis and immortalisation. *Genes Dev.* **12**, 2424–2433 (1998).
26. de Stanchina, E. *et al.* E1A signalling to p53 involves the p19^{Arf} tumour suppressor. *Genes Dev.* **12**, 2434–2442 (1998).
27. Blyth, K. *et al.* Synergy between a human *c-myc* transgene and p53 null genotype in murine thymic lymphomas: contrasting effects of homozygous and heterozygous p53 loss. *Oncogene* **10**, 1717–1723 (1995).
28. Bodrug, S. E. *et al.* Cyclin D1 transgene impedes lymphocyte maturation and collaborates in lymphomagenesis with the *myc* gene. *EMBO J.* **13**, 2124–2130 (1994).
29. Mateyak, M. K., Obaya, A. J., Adachi, S. & Sedivy, J. M. Phenotypes of *c-Myc*-deficient Rat fibroblasts isolated by targeted homologous recombination. *Cell Growth Differ.* **8**, 1039–1048 (1997).
30. Hall, M. & Peters, G. Genetic alterations of cyclins, cyclin-dependent kinases and cdk inhibitors in human cancer. *Adv. Cancer Res.* **68**, 67–108 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank G. Peters for p16 and p15 cDNA plasmids; H. Koseki and J. Deschamps for *mel-18*^{-/-} embryos; T. Ide for the TiG-3 cells; J. Sedivy for *c-myc*^{-/-} TGR-1 cells; S. Lowe, B. Amati, C. Sherr, M. Ewen, D. Peiper and R. Bernards for gifts of recombinant retroviral vectors; G. Nolan for providing phoenix eco- and amphotropic packaging cell lines and LZRS-IRES-EGFP retroviral vectors; N. van der Lugt for constructing the LZRS-*bmi-1*-IRES-EGFP retrovirus; E. de Pauw for telomere FISH analysis; and R. Bernards, A. Berns and W. Voncken for critically reading the manuscript J.J.L.J. and K.K. were supported by a grant of the Dutch Cancer Society (K.W.F.).

Correspondence and requests for materials should be addressed to M.v.L. (e-mail: lohuiizen@nki.nl).

Robustness in bacterial chemotaxis

U. Alon^{*†}, M. G. Surette[‡], N. Barkai[†] & S. Leibler^{*†}

Departments of ^{*}Molecular Biology and [†]Physics, Princeton University, Princeton, New Jersey 08544, USA

[‡] Department of Microbiology and Infectious Diseases, Calgary, Alberta, Canada T2N 4N1

Networks of interacting proteins orchestrate the responses of living cells to a variety of external stimuli¹, but how sensitive is the functioning of these protein networks to variations in their biochemical parameters? One possibility is that to achieve appropriate function, the reaction rate constants and enzyme concentrations need to be adjusted in a precise manner, and any deviation from these 'fine-tuned' values ruins the network's performance. An alternative possibility is that key properties of

biochemical networks are robust²; that is, they are insensitive to the precise values of the biochemical parameters. Here we address this issue in experiments using chemotaxis of *Escherichia coli*, one of the best-characterized sensory systems^{3,4}. We focus on how response and adaptation to attractant signals vary with systematic changes in the intracellular concentration of the components of the chemotaxis network. We find that some properties, such as steady-state behaviour and adaptation time, show strong variations in response to varying protein concentrations. In contrast, the precision of adaptation is robust and does not vary with the protein concentrations. This is consistent with a recently proposed molecular mechanism for exact adaptation, where robustness is a direct consequence of the network's architecture².

The *E. coli* chemotaxis system^{3,4} has emerged as a prototype for understanding how processes at the network level arise from interactions between individual components^{5,6}. The protein net-

work responsible for chemotaxis has been characterized in detail^{3,4} (Box 1). This sensory network governs the migration of bacteria towards chemical attractants and away from repellents by translating temporal changes in the level of chemical stimuli into a modulation of the cell's swimming direction. This is achieved by controlling the frequency of abrupt direction changes called tumbles. An important feature of chemotaxis is exact adaptation: a change in the concentration of a chemical stimulant induces a rapid change in the bacteria's tumbling frequency, which gradually adapts back precisely to its pre-stimulus value^{7,8}.

Once the components of a biochemical network are isolated and their interactions characterized, the mechanisms of the network's functioning can be addressed. Here, we focus on a distinct, complementary aspect, namely the sensitivity of the network's functioning to variations in its biochemical parameters. Specifically, we ask how sensitive exact adaptation in chemotaxis is to variations in the concentration of proteins in the network. Several theoretical proposals for the molecular mechanism of adaptation^{5,9–11} required that the biochemical parameters of the chemotaxis network have to be fine-tuned to obtain exact adaptation. Deviations of the parameters from their fine-tuned values would result only in partial adaptation. In contrast, a recent analysis², based on a two-state model of chemoreceptors^{3,12}, proposed that exact adaptation is a robust property of chemotaxis. That is, the ability of the bacteria to adapt precisely would survive substantial variations in any of the biochemical parameters of the network.

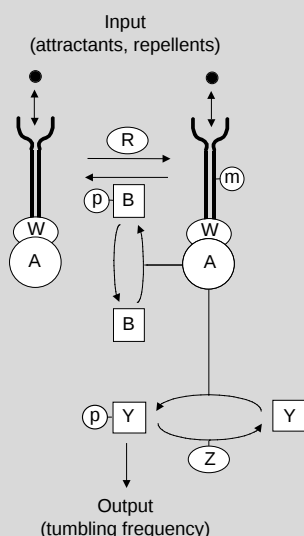
To differentiate between the fine-tuned and robust scenarios, we systematically varied the concentrations of the chemotaxis-network proteins, and measured the resulting behaviour. Similar methods have been used in analyses of metabolic pathways¹³. The tumbling frequency, averaged over a population of swimming cells, was measured using a video microscopy image-analysis system¹⁴. Cells were stimulated by addition of a saturating amount of attractant (1 mM L-aspartate), and their behaviour was compared with that of unstimulated cells (Fig. 1). The cells responded to the attractant by suppressing tumbling. The tumbling frequency then increased as the cells adapted, with a characteristic adaptation time, reaching a new steady state. The adaptation precision, P , was defined as the ratio between the steady-state tumbling frequency of unstimulated and stimulated cells: wild-type cells showed exact adaptation^{7,8,15} ($P = 0.98 \pm 0.05$, mean $\pm$ standard deviation (s.d.)). The same degree of adaptation precision was also found for other attractants and repellents (1 mM L-serine, 50 mM L-leucine, data not shown. Note, however, that a lower steady-state tumbling frequency was reported¹⁵ in the presence of L-serine than in its absence under different buffer conditions).

One of the key proteins in the adaptation mechanism^{3,4} is CheR, which is responsible for chemoreceptor methylation (Box 1). A strain deleted for *cheR* did not tumble and could not adapt (Fig. 2). Tumbling was restored by CheR expression under a *lac* promoter from a low-copy plasmid. Intracellular CheR concentrations were varied, over a range of ~ 100 -fold, by induction with different amounts of isopropyl- β -D-thiogalactoside (IPTG), and the average CheR concentration was quantitated by immunoblots. The initial smooth swimming response of the cells to attractant (< 0.05 tumbles per s) was followed by adaptation. As CheR was varied between ~ 0.4 and 6 times wild-type concentration, the adaptation time, τ , varied more than 20-fold, from 23 ± 2 min to ~ 1 min (Fig. 2b). Similarly, the steady-state tumbling frequency, f , increased by about threefold. This is consistent qualitatively with model predictions² where $\tau \sim 1/[\text{CheR}]$ and f increases with the intracellular CheR concentration $[\text{CheR}]$. The tumbling frequency saturated at about 0.75 tumbles per s at the highest concentration of CheR attained (about 50 times the wild-type concentration). Throughout these variations, adaptation remained precise to within experimental error (Fig. 2a).

Similar results were obtained when the concentrations of the

Box 1 Chemotaxis system of *E. coli*

Bacteria such as *E. coli* bias their swimming motion towards specific attractants and away from repellents^{3,4}. Bacterial motion resembles a random walk, with periods of smooth swimming interrupted by brief tumbles that change the swimming direction¹⁵. Chemotaxis is achieved by modulation of the tumbling frequency. When moving up an attractant gradient, the bacteria encounter an attractant concentration that increases with time. In response, they tumble less frequently and thus tend to continue to move up the gradient. This process is mediated by a protein network^{3,4} (see Box Figure below).



Information about the chemical environment is transduced into the cells by chemoreceptors, such as the aspartate receptor Tar, which span the membrane. The chemoreceptors form complexes inside the cells with the kinase CheA (A) and CheW (W). CheA phosphorylates itself and then transfers phosphoryl (P) groups to CheY (Y), a diffusible messenger protein. The phosphorylated form of CheY interacts with the flagellar motors to induce tumbles. The rate of CheY dephosphorylation is greatly enhanced by CheZ (Z). Binding of attractants to the receptors decreases the rate of CheY phosphorylation and tumbling is reduced. Adaptation is provided by changes in the level of methylation of the chemoreceptors: methylation increases the rate of CheY phosphorylation. A pair of enzymes, CheR (R) and CheB (B), add and remove methyl (m) groups. To adapt to an attractant, methylation of the receptors must rise to overcome the suppression of receptor activity caused by the attractant binding. CheA enhances the demethylating activity of CheB by phosphorylating CheB on its amino-terminal domain^{20,21}.

other proteins in the network were varied (Table 1). The steady-state tumbling frequency was affected by changes in concentration of these proteins, in qualitative agreement with previous studies^{16–19}. For instance, an increase in the concentration of CheB, the receptor-demethylating enzyme (Box 1), caused a decrease in tumbling frequency and an increase in adaptation time. Loss of CheZ resulted in increased tumbling frequency, and the initial response to stimulation was a reduction in the tumbling frequency to about half of its adapted value, rather than to <0.05 tumbles per s as in the wild-type strain. This can be explained by the greatly reduced rate of CheY dephosphorylation in the absence of CheZ. Both tumbling frequency and adaptation time were modified upon simultaneous overexpression of the entire *meche* operon (encoding Tar, Tap, CheR, CheB, CheY and CheZ) from a plasmid. Strikingly, exact adaptation was observed for all of these perturbed networks (Table 1), which is consistent with the predictions of the robust-adaptation model².

To probe the mechanism for exact adaptation further, we used an activated mutant of CheB, CheBc, which can demethylate the receptors but lacks the domain phosphorylated by CheA^{20,21}. This mutant lacks the feedback loop that was often thought to be responsible for exact adaptation^{4,10,11}. In this feedback loop, the receptor-controlled kinase, CheA, phosphorylates CheB, increasing the demethylation of the receptors and thereby downregulating its own activity (Box 1). We find that cells deleted for *CheB* on the chromosome, but which express CheBc, show response and exact adaptation to attractant stimuli (Table 1). This seems to exclude CheB phosphorylation as the mechanism of exact adaptation, although it may be important in other aspects of chemotaxis^{19,20}. In this context, the postulated molecular mechanism that leads to robust exact adaptation, analysed in ref. 2, is not based on the phosphorylation feedback loop. The crux of this mechanism is a different type of feedback in which CheB demethylates receptors only when they are in their active conformation^{2,3,12}.

The present results show that exact adaptation is maintained despite substantial variations in the network-protein concentra-

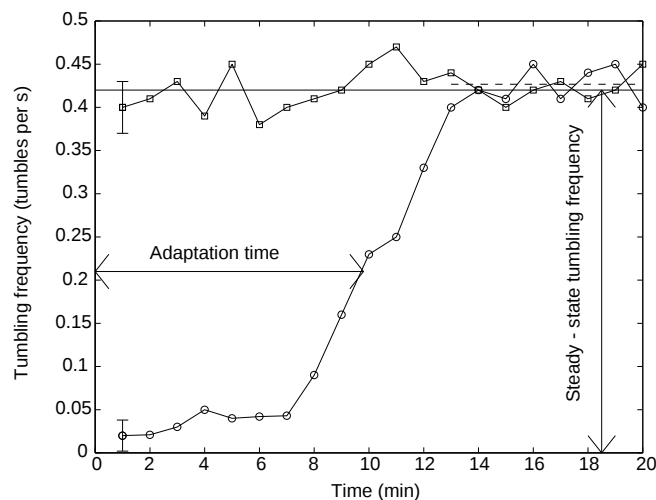


Figure 1 Tumbling frequency as a function of time for wild-type (RP437) cells. Circles: cells stimulated at time $t = 0$ by mixing with saturating attractant (1 mM L-aspartate). Squares: unstimulated cells (mock-mixed with chemotaxis buffer). Tumbling frequency was determined using computerized video tracking¹⁴. Each point represents data from 10 s motion of 100–400 cells. The adaptation time was defined as the time where the tumbling frequency of stimulated cells rises to halfway between its earliest measured value and its steady-state value. Precision of adaptation was defined as the ratio between the steady-state tumbling frequency of unstimulated cells (full horizontal line) and stimulated cells (dashed horizontal line).

tions. Why is adaptation precision a robust property, whereas properties such as adaptation time and steady-state behaviour are sensitive to variations in biochemical parameters? It is tempting to argue that properties that are critical to the functioning of the network are selected to be robust so that they can withstand natural variations. There is indeed some evidence that exact adaptation is important in the sensory process. It allows the system to compensate for the presence of continued stimulation, and to be ready to respond to further stimuli. Mutant bacteria that show only partial adaptation (strains deleted for both *cheR* and *cheB*^{22–24}) are severely deficient in chemotaxis, as assayed by their net motion in attractant gradients^{24,25}, despite the fact that their steady-state tumbling frequency is close to that of the wild-type strain. In contrast, strains that have normal adaptation, but tumbling frequencies different from wild type, show chemotaxis ability that is comparable to the wild-type strain²⁶. We confirmed this result using the present strains with varying CheR concentrations (data not shown). Furthermore, exact adaptation is found in other bacteria such as *Bacillus subtilis*²⁷ and *Rhodobacter sphaeroides*²⁸. These results suggest that exact adaptation is critical for chemotaxis, whereas chemotaxis ability is not dependent on the precise value of the steady-state tumbling

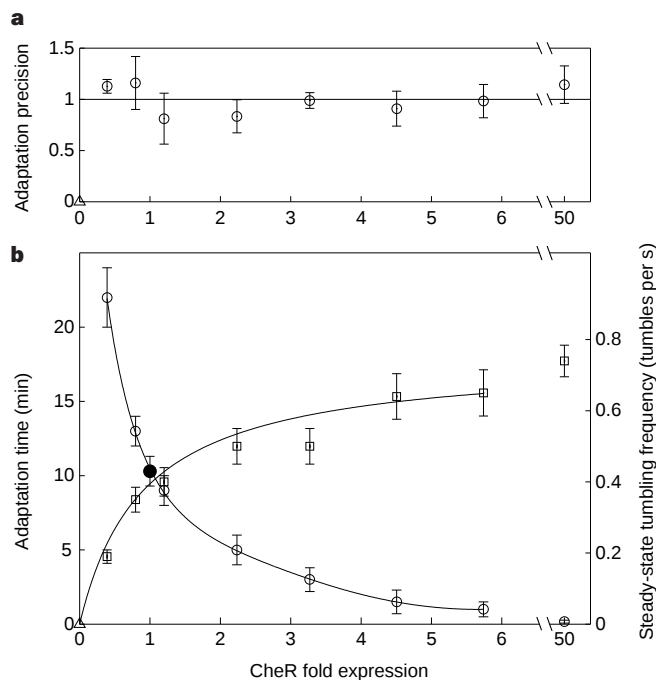


Figure 2 Chemotaxis behaviour of cells with varying intracellular concentration of the protein CheR. CheR was expressed from the plasmid pUA4 with varying levels of IPTG induction in a strain deleted for *cheR* (RP4968). **a**, Precision of adaptation, defined as the ratio between the steady-state tumbling frequency of unstimulated cells and cells stimulated with 1 mM L-aspartate. A precision of 1.0 corresponds to exact adaptation. Cells lacking CheR (RP4968 with the control vector pHSG575, triangle) did not respond to attractants, but showed a persistent response of about 0.6 tumbles per s to repellent (50 mM L-leucine). **b**, Average steady-state tumbling frequency of unstimulated cells (open squares, right scale), and average adaptation time to a step-like stimulation with 1 mM L-aspartate (open circles, left scale). Solid circle, wild-type strain (RP437 + pHSG575). Triangle, tumbling frequency of RP4968 + pHSG575. Lines are guides to the eye. Relative CheR expression was measured by immunoblots. 'Wild-type' CheR concentration was defined as the induction level where the adaptation time was equal to that of RP437 + pHSG575. Immunoblots also showed that the level of other chemotaxis proteins (CheB and CheY) did not vary measurably with CheR expression. Errors in relative CheR level are estimated to be under 30%. Mean and standard deviation of triplicate experiments are shown.

Table 1 Effect of variation in chemotaxis protein levels on behaviour

Protein varied	Fold expression	Strain background	Steady-state tumbling frequency (s ⁻¹)	Adaptation time (min)	Precision of adaptation
Wild type	1.0	Wild type (RP437)	0.44 ± 0.03	10 ± 1	0.98 ± 0.05
CheB*	0.4 ± 0.1	ΔcheB (RP4972)	0.66 ± 0.05	7 ± 1	0.98 ± 0.12
CheB*	12 ± 3	ΔcheB (RP4972)	0.14 ± 0.02	15 ± 1	1.09 ± 0.11
CheBc†	~11	ΔcheB (RP4972)	0.74 ± 0.06	9 ± 2	0.90 ± 0.13
CheY‡	0.2 ± 0.1	ΔcheY _Z (RP5231)	0.24 ± 0.04	11 ± 3	1.04 ± 0.08
CheZ	0	ΔcheZ (RP1616)	1.6 ± 0.1	10 ± 2	1.1 ± 0.14
Tar, Tap, CheR,B,Y,Z§	5 ± 2¶	Wild-type (RP437)	0.30 ± 0.06	3 ± 1	1.04 ± 0.07

Cells were stimulated with 1 mM L-aspartate. Behavioural results are mean and standard deviation of triplicate experiments. Fold expression is relative to RP437, shown as mean and standard deviation of four independent immunoblot determinations using affinity-purified polyclonal antibodies.

* Plasmid pUA3.

† Plasmid pME304 (ref. 21).

‡ Plasmid pLC576 (ref. 14).

§ Plasmid pCR73 (ref. 18).

¶ CheBc fold expression estimated against wild-type CheB in RP437.

¶ Determined using antibodies against CheR and CheB.

frequency and the adaptation time. Exact adaptation might also be a 'side effect' of the mechanism for a different, critical aspect of chemotaxis (for example, the amplification or gain of the system^{6,22}), and as such could be selected as a robust property.

The experimental approach presented here can be generalized to other biochemical or genetic networks. For a given network, those functional properties that are robust, and those that are sensitive with respect to parameter variations could be classified. It would then be interesting to see whether any 'design principles' connected with robustness, perhaps analogous to those used in engineering, emerge from such systematic studies. □

Methods

Bacterial strains and plasmids. All strains were derivatives of *E. coli* K12 strain RP437 [*thr leu his metF eda rpsL thi ara lacY xyl tonA tsx*], which is wild type for chemotaxis¹⁶. pUA4 was constructed by subcloning an *XbaI*–*HindIII* fragment of pME1 (ref. 29) containing *cheR* from *S. typhimurium* and the *EcoRI*–*SbaI* fragment of pUC054 (ref. 30) containing *lacI^q* into the low-copy vector pHSG575 (ref. 14). An *EcoRI*–*HindIII* fragment of pME30 (ref. 29) containing *cheB* from *S. typhimurium* was subcloned into pBAD18 (ref. 14) to yield pUA3. Similar results were obtained with cells expressing *E. coli* CheR and CheB from the plasmids p43:cheR (ref. 18) and p43:cheB (ref. 18). Other plasmids are listed in Table 1.

Behavioural assays. Cultures, grown overnight at 30 °C in Tryptone broth¹⁴ with selective antibiotics, were diluted 1:50 into 20 ml Tryptone broth in 125-ml flasks, shaken at 200 r.p.m. at 30 °C, induced after 2 h and collected at $A_{600} = 0.5$ (5×10^8 cells ml⁻¹). Cells were washed twice at 800g into chemotaxis buffer¹⁴ (7.6 mM (NH₄)₂, 2 mM MgSO₄, 20 μM FeSO₄, 0.1 mM EDTA, 0.1 mM L-methionine, 60 mM potassium phosphate, pH 6.8), and incubated for 15 min at room temperature. A sample (10 μl) of cell suspension was stimulated by mixing with 10 μl of 2 mM L-aspartate in chemotaxis buffer, or mock stimulated by mixing with 10 μl chemotaxis buffer. Cells (0.5 μl) were then placed in the centre of a circle inscribed with a china marker on a vinyl slide and covered by a vinyl coverslip (Fisher), creating a several-micron-thick fluid layer. To prevent cell adhesion to the surfaces, slides and coverslips were precoated by dipping in a solution of 0.1% bovine serum albumin (BSA) (>99% purity; Sigma) in chemotaxis buffer followed by a brief wash in distilled water. Bacteria were observed with dark-field video microscopy (Nikon Optiphot-2; field of view spanned 220 μm), and average tumbling frequency was determined by computerized image analysis as previously described¹⁴. The behaviour with 1 mM L-aspartate stimuli was indistinguishable from that observed in control experiments with 0.5 mM or 10 mM L-aspartate stimuli. Chemotaxis ability was assayed on soft agar plates (Tryptone broth + 0.3% agar) at 30 °C¹⁷.

Protein quantitation. Cultures, grown as for the behavioural assays, were subjected to SDS–polyacrylamide gel electrophoresis and western blot analysis as described¹⁴. To determine relative expression in strains A and B, where B has a higher copy number per cell of the protein assayed, lysates of strain B were diluted with various amounts of PS2002 (ref. 14) lysate (an RP437 derivative deleted for *Tar*, *Tap* and all of the *che* genes) and compared with standards of

strain A (equal total protein in each lane). The diluted strain intensities bracketed the standard intensity.

Received 13 October; accepted 27 November 1998.

- Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307–312 (1995).
- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Stock, J. B. & Surette, M. G. in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1103–1129 (ASM Press, Washington, 1996).
- Falke, J. J., Bass, R. B., Butler, S. L., Chervitz, S. A. & Danielson, M. A. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.* **13**, 457–512 (1997).
- Bray, D., Bourret, R. B. & Simon, M. I. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **5**, 469–482 (1993).
- Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
- Macnab, R. M. & Koshland, D. E. The gradient sensing mechanism in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **69**, 2509–2512 (1972).
- Berg, H. C. & Tedesco, P. Transient response to chemotaxis stimuli in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **72**, 3235–3239 (1975).
- Segel, L. A., Goldbeter, A., Devrotes, P. N. & Knox, B. E. A mechanism for exact sensory adaptation based on receptor modification. *J. Theor. Biol.* **120**, 151–179 (1986).
- Hauri, D. C. & Ross, J. A. A model of excitation and adaptation in bacterial chemotaxis. *Biophys. J.* **68**, 708–722 (1995).
- Spiro, P. A., Parkinson, J. S. & Othmer, H. G. A model of excitation and adaptation in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **94**, 7263–7268 (1997).
- Asakura, S. & Honda, H. Two-state model for bacterial chemoreceptor proteins. *J. Mol. Biol.* **176**, 349–367 (1984).
- Fell, D. *Understanding the Control of Metabolism* (Portland Press, London, 1997).
- Alon, U. et al. Response regulator output in bacterial chemotaxis. *EMBO J.* **17**, 4238–4248 (1998).
- Berg, H. C. & Brown, D. A. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**, 500–504 (1972).
- Parkinson, J. S. & Houts, S. Isolation and behavior of *Escherichia coli* deletion mutants lacking chemotaxis function. *J. Bacteriol.* **151**, 106–113 (1982).
- Wolfe, A. J., Conley, P. M., Kramer, T. J. & Berg, H. C. Reconstitution of signaling in bacterial chemotaxis. *J. Bacteriol.* **169**, 1878–1885 (1987).
- Russel, C. B., Stewart, R. C. & Dahlquist, F. W. Control of transducer methylation levels in *Escherichia coli*: investigation of components essential for modulation of methylation and demethylation reactions. *J. Bacteriol.* **171**, 3609–3618 (1989).
- Levin, M., Morton-Firth, C., Abouhamad, W., Bourret, R. & Bray, D. Origins of individual swimming behavior in bacteria. *Biophys. J.* **74**, 175–181 (1998).
- Stewart, R. C., Russel, C. B., Roth, A. F. & Dahlquist, F. W. Interaction of CheB with chemotaxis signal transduction components in *Escherichia coli*: modulation of the methyltransferase activity and effects on cell swimming behavior. *Cold Spring Harb. Symp. Quant. Biol. LIII*, 27–40 (1988).
- Lupas, A. & Stock, J. B. Phosphorylation of an N-terminus regulatory domain activates the CheB methyltransferase in bacterial chemotaxis. *J. Biol. Chem.* **264**, 17337–17342 (1989).
- Segall, J. E., Block, S. M. & Berg, H. C. Temporal comparisons in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **83**, 8987–8991 (1986).
- Stock, J., Kersulis, G. & Koshland, D. E. Neither methylating nor demethylating enzymes are required for bacterial chemotaxis. *Cell* **42**, 683–690 (1985).
- Weis, R. M. & Koshland, D. E. Reversible methylation is essential for normal chemotaxis in *Escherichia coli* in gradients of aspartic acid. *Proc. Natl Acad. Sci. USA* **86**, 83–87 (1989).
- Berg, H. & Turner, L. Chemotaxis of bacteria in glass capillary arrays. *Biophys. J.* **58**, 919–930 (1990).
- Weis, R. M. & Koshland, D. E. Chemotaxis in *Escherichia coli* proceeds efficiently from different initial tumble frequencies. *J. Bacteriol.* **172**, 1099–1105 (1990).
- Kirsch, M. L. et al. Chemotactic methyltransferase promotes adaptation to repellents in *Bacillus subtilis*. *J. Biol. Chem.* **268**, 25350–25356 (1993).
- Grishanin, R. N., Gauden, D. E. & Armitage, J. P. Photoresponses in *Rhodospirillum rubrum*: role of photosynthetic electron transport. *J. Bacteriol.* **179**, 24–30 (1997).
- Simm, A. S., Keane, M. G. & Stock, J. B. Multiple forms of the CheB methyltransferase in bacterial chemotaxis. *J. Biol. Chem.* **260**, 10161–10168 (1985).
- Surette, M. G. & Stock, J. B. Role of α-helical coiled-coil interactions in receptor dimerization, signaling and adaptation during bacterial chemotaxis. *J. Biol. Chem.* **271**, 17966–17973 (1996).

Acknowledgements. We thank S. Block, H. Berg, P. Cluzel, M. Elowitz, L. Hartwell, M. Levitt, A. Murray, M. Simon, J. Staropoli, R. Stewart and J. Stock for discussions, and J. Parkinson, F. Dahlquist and J. Stock for providing strains, antibodies and plasmids. N.B. is a Dickel fellow. U.A. is a Rothchild and a Markee fellow; he dedicates this work to the memory of I. Benton.

Correspondence and requests for materials should be addressed to S.L. (e-mail: leibler@princeton.edu).

Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein

Antonia Lopez-Girona*, Beth Furnari*, Odile Mondesert & Paul Russell

Departments of Molecular Biology and Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA

* These authors contributed equally to this work

DNA damage activates a cell-cycle checkpoint that prevents mitosis while DNA repair is under way¹. The protein Chk1 enforces this checkpoint by phosphorylating the mitotic inducer Cdc25 (refs 2–6). Phosphorylation of Cdc25 by Chk1 creates a binding site in Cdc25 for 14-3-3 proteins^{5–8}, but it is not known how 14-3-3 proteins regulate Cdc25. Rad24 is a 14-3-3 protein that is important in the DNA-damage checkpoint in fission yeast⁹. Here we show that Rad24 controls the intracellular distribution of Cdc25. Elimination of Rad24 causes nuclear accumulation of Cdc25. Activation of the DNA-damage checkpoint causes the net nuclear export of Cdc25 by a process that requires Chk1, Rad24 and nuclear-export machinery. Mutation of a putative nuclear-export signal in Rad24 impairs the nuclear exclusion of Rad24, the damage-induced nuclear export of Cdc25 and the damage check-

point. Thus, Rad24 appears to function as an attachable nuclear-export signal that enhances the nuclear export of Cdc25 in response to DNA damage.

Mitotic events are initiated by the cyclin-dependent kinase Cdc2. The protein phosphatase Cdc25 activates Cdc2 by dephosphorylating residue Y15 of Cdc2. In the fission yeast *Schizosaccharomyces pombe*, Cdc2 is localized in the nucleus before mitosis¹⁰. The nuclear membrane remains intact throughout the cell cycle in fission yeast; thus Cdc25 must presumably enter the nucleus to activate Cdc2. We studied this possibility by monitoring Cdc25 localization during the cell cycle in a strain that expressed Myc-tagged Cdc25 from the *cdc25*⁺ genomic locus (Fig. 1a). We detected Cdc25 in both the cytoplasm and the nucleus in cells in the G2 phase of the cell cycle, that is, before mitosis, with a higher concentration of Cdc25 in the nucleus (Fig. 1a). However, the nucleus occupies a small fraction of the cell volume, so most of the Cdc25 is apparently in the cytoplasm. This conclusion was supported by the results of cell-fractionation studies (see below). The amount of nuclear Cdc25 appeared to increase in late G2 phase and decreased during anaphase (Fig. 1a). The amount of Cdc25 in the nucleus was reduced substantially in binucleate and septated cells (Fig. 1b). These cells were in late M (mitosis), G1 or S (DNA-replication) phase. Degradation of Cdc25 might contribute to the decrease of the Cdc25 nuclear signal in late M phase¹¹. Quantitative analysis of signal intensity confirmed that nuclear accumulation of Cdc25 was enhanced in late G2 relative to early G2 or G1/S phase (Fig. 1c).

The DNA-damage checkpoint prevents Cdc25 from dephosphorylating Cdc2 on Y15 *in vivo*^{3,4}. Chk1 might inhibit Cdc25 or reduce

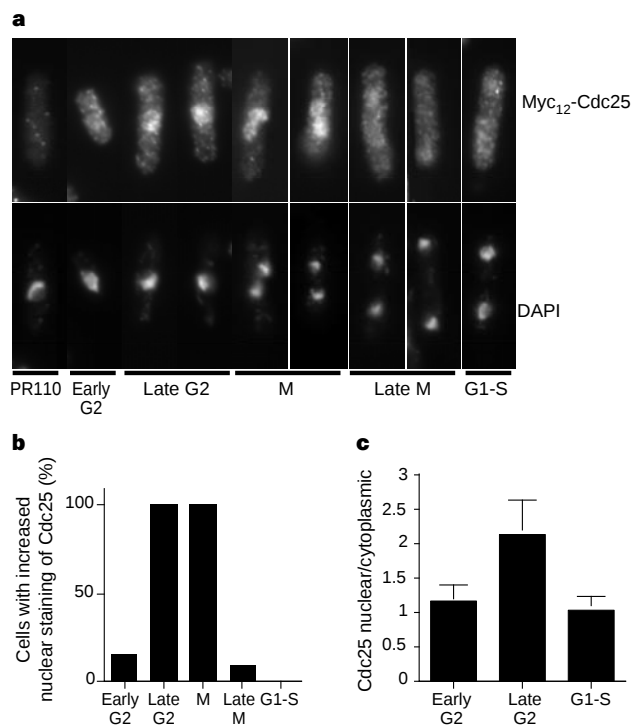


Figure 1 Cdc25 nuclear staining is increased in late G2 and M phases. **a**, Asynchronous OM1730 (*cdc25:12myc*) cells were processed for indirect immunofluorescence microscopy using anti-Myc antibodies. Nuclei were stained with DAPI. Approximate phase of the cell cycle is indicated (see Methods). The Cdc25 nuclear signal was highest in late G2 and lowest in late M, G1 or S phases. PR110 is the negative control. **b**, Cells at different stages in the cell cycle were scored for enhanced nuclear staining of Cdc25. **c**, The nuclear/cytoplasmic ratio of Cdc25 signal was quantified in early G2, late G2 and G1/S cells.

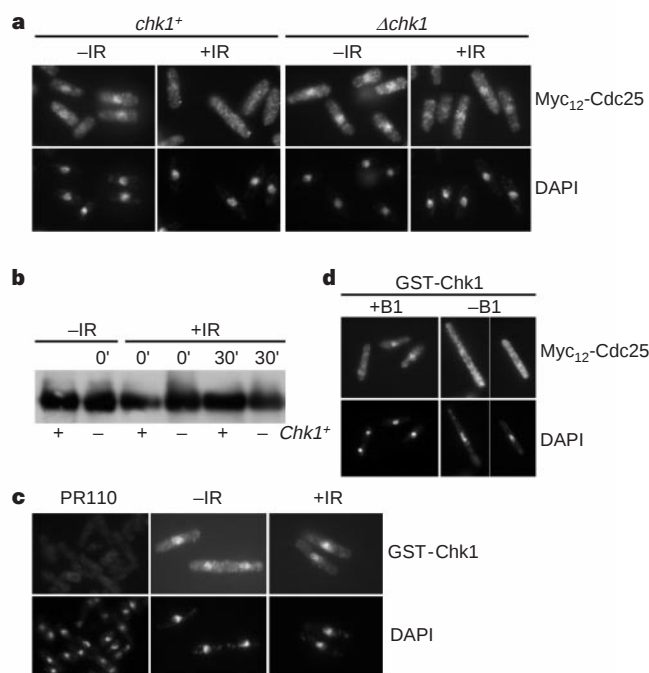


Figure 2 Checkpoint-induced depletion of nuclear Cdc25. **a**, Irradiation causes Chk1-dependent nuclear depletion of Cdc25. OM1730 (*cdc25:12myc*) and BF1903 (*cdc25:12myc Δchk1*) cells were irradiated (+IR) or mock-irradiated (–IR) and processed to detect Cdc25:12myc. Nuclei were stained with DAPI. **b**, Cdc25 abundance is unchanged by irradiation. OM1715 (*cdc25:6HA*) and BF1907 (*cdc25:6HA Δchk1*; HA is haemagglutinin) were irradiated for 0 or 30 min or mock-irradiated. Immunoblot analysis detected HA₆-Cdc25. **c**, Chk1 is nuclear. GST-Chk1 was localized in BF2220 (*nmt1:GST:chk1*) cells grown in medium that represses the *nmt1* promoter and thus approximates wild-type Chk1 expression. PR110 is the negative control strain. **d**, Chk1 overexpression induces nuclear depletion of Cdc25. AL2277 (*cdc25:12myc nmt1:GST:chk1*) cells were grown in *nmt1*-repressing (+B1) or -inducing (–B1) medium.

access to Cdc2, perhaps by regulating Cdc25 localization. We explored the latter possibility by exposing cells to ionizing radiation. In wild-type cells, ionizing radiation caused a reduction in the amount of nuclear Cdc25 (Fig. 2a). In contrast, ionizing radiation did not change Cdc25 localization in $\Delta chk1$ cells (Fig. 2a): nuclear accumulation of Cdc25 was detected in 68% of the irradiated $\Delta chk1$ cells as compared with 8% of wild-type cells. Immunoblot analysis of a separate experiment showed that the total amount of Cdc25 was unchanged in irradiated wild-type and $\Delta chk1$ cells (Fig. 2b), indicating that ionizing radiation caused net nuclear export of Cdc25. We detected Chk1 as a nuclear protein when it was expressed at low levels as a fusion protein with glutathione-S-transferase (GST) (Fig. 2c). High expression of GST-Chk1 induced cell-cycle arrest, mimicking a damage checkpoint, and caused nuclear exclusion of Cdc25 (Fig. 2d). These data indicate that Chk1 may phosphorylate the nuclear fraction of Cdc25 and promote nuclear export of Cdc25.

Phosphorylation of human Cdc25C by Chk1 creates a binding site for 14-3-3 proteins^{5,6}. Rad24, a 14-3-3 protein in fission yeast, is involved in cell-cycle and checkpoint controls⁹. A *rad24* mutant undergoes mitosis at a reduced cell length, is sensitive to DNA damage and exhibits a substantially attenuated checkpoint arrest when exposed to γ -irradiation⁹. We observed that the Cdc25 nuclear signal was more intense in *rad24* mutant cells than in wild-type cells (Fig. 3a). This fact was particularly evident from comparison of binucleate and septated cells: nuclear accumulation of Cdc25 was detected in ~82% (18 of 22) of binucleate and septated *rad24* cells and in ~3% (1 of 30) of wild-type cells. These results were confirmed in cell-fractionation experiments, which indicated that most Cdc25 was cytosolic in wild-type cells but nuclear in *rad24* cells (Fig. 3b). In *rad24* cells, the amount of Cdc25 in the nucleus remained high following irradiation (Fig. 3a). These results show

that Rad24 is required for proper localization of Cdc25 in normal growth conditions and in response to DNA damage.

Rad24 might have a direct role in regulating Cdc25 localization, perhaps by binding to Cdc25 and escorting it from the nucleus. This possibility was explored by determining the localization of Rad24. We replaced genomic *rad24* with a construct encoding Rad24 tagged with the haemagglutinin epitope. Immunolocalization analysis showed that Rad24 was excluded from the nucleus in ~95% (21 of 22) of mock-irradiated wild-type cells (Fig. 3c). This pattern of Rad24 localization was unchanged in response to ionizing radiation; Rad24 was excluded from nucleus in ~97% (34 of 35) of cells.

The association between Rad24 and GST-Cdc25 was undiminished in $\Delta chk1$ cells (Fig. 3d). This finding was confirmed with immunoprecipitation studies of Cdc25 (Fig. 3e). 14-3-3 proteins bind to phosphorylated motifs^{12,13}; thus there are apparently other kinases, in addition to Chk1, that phosphorylate Cdc25 at 14-3-3-binding sites. These kinases might include Cds1 in fission yeast and C-TAK1 in mammalian cells^{8,14}. The amount of Cdc25 associated with Rad24 was not substantially increased in response to ionizing radiation (Fig. 3e), although a modest increase cannot be excluded. These findings indicate that most Cdc25 may be bound to 14-3-3 proteins in the absence of DNA damage. Similar conclusions have been made from studies of Cdc25 in mammalian cells and *Xenopus* egg extracts^{6,15}. These results indicate that Chk1 may target the nuclear pool of Cdc25 that is presumably unbound to Rad24.

We determined whether damage-induced nuclear depletion of Cdc25 required Crm1, a highly conserved nuclear-export factor¹⁶⁻¹⁹. At the semi-permissive temperature of 30 °C, the cold-sensitive *crm1-809* mutant is partially deficient in nuclear export¹⁷. Under these conditions, *crm1-809* cells showed a marked nuclear accumulation of the Cdc25 (Fig. 4a). Binucleate and septated *crm1-809* cells

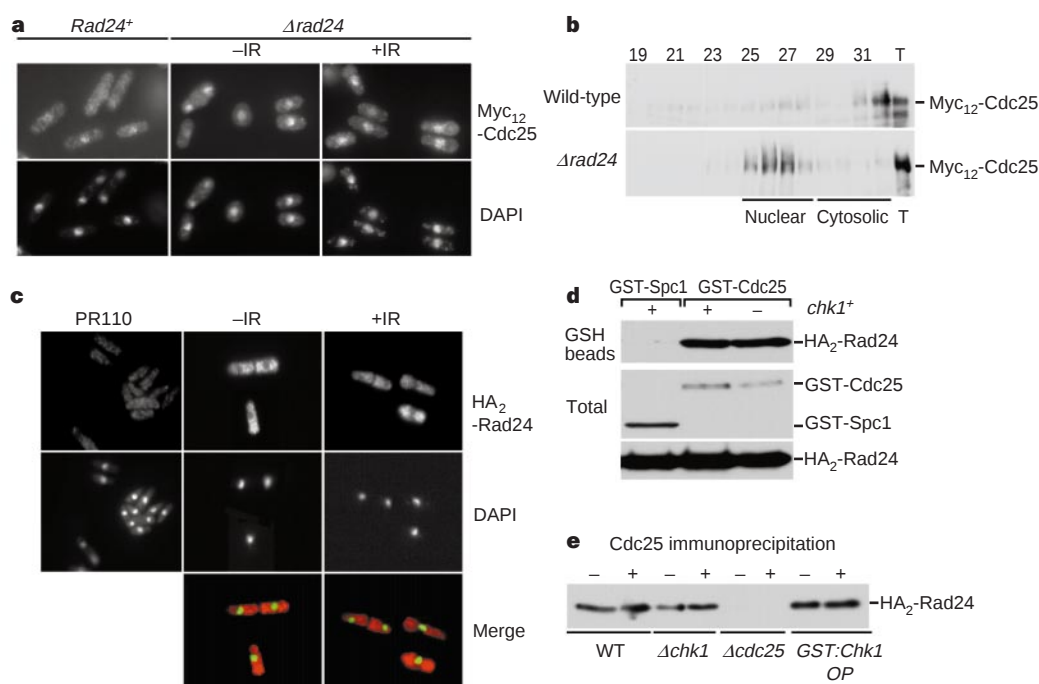


Figure 3 Rad24 regulates Cdc25 localization. **a**, Cdc25 nuclear localization is enhanced in $\Delta rad24$ cells in the presence or absence of ionizing radiation (IR). Nuclei were stained with DAPI. **b**, Cellular fractionation indicates that Cdc25 is predominantly cytoplasmic in wild-type cells but nuclear in $\Delta rad24$ cells. Numbers above lanes indicate cellular fractions. T, total extracts. **c**, Pseudocolour merged image of stained HA₂-Rad24 (red) and nuclei (green) shows that Rad24 is excluded from the nucleus in the presence or absence of ionizing irradiation.

PR110 is the negative control strain. **d**, HA₂-Rad24 associates with GST-Cdc25 in a Chk1-independent manner. GST-Cdc25 or GST-Spc1 were purified with glutathione-Sepharose (GSH) beads and processed to detect HA₂-Rad24. Bottom, HA₂-Rad was used as a loading control. **e**, Cdc25 associates with Rad24 independently of irradiation or Chk1 expression. Strains with the indicated genotypes were irradiated (+) or mock-irradiated (-) and HA₂-Rad24 was detected in Cdc25 immunoprecipitations.

exhibited an intense Cdc25 signal in the nucleus, unlike wild-type cells (Fig. 4a). In *crm1-809* cells, nuclear accumulation of Cdc25 was undiminished in response to ionizing radiation (Fig. 4a). The *crm1-809* cells were moderately elongated relative to wild-type, indicating that a general defect in nuclear exclusion of proteins may cause a mitotic delay that is not overcome by increased nuclear localization of Cdc25. However, *crm1-809* cells grown at 30 °C exhibit increased sensitivity to ultraviolet and γ -irradiation (data not shown), at a level approximately equivalent to the sensitivity of *rad24* mutants, supporting the hypothesis that nuclear export is important for the DNA-damage response.

These results indicated that Crm1 may be required for nuclear export of Cdc25. Crm1 associates with proteins that contain the leucine-rich nuclear-export signal (NES) sequence. Cdc25 contains no obvious matches to the NES consensus sequence¹⁷. However, Rad24 has a potential NES in the region of amino acids 219–234 (Fig. 4b). Indeed, immunolocalization studies showed that nuclear exclusion of Rad24 was substantially diminished in *crm1-809* cells (Fig. 4c): we detected nuclear exclusion of Rad24 in ~100% (25 of 25) of wild-type cells as compared with ~22% (4 of 18) of the *crm1-809* cells. We replaced the genomic copy of *rad24*⁺ with a mutant gene, *rad24-nes*, encoding a protein in which I222 and L226 were each replaced with alanine. Equivalent mutations disrupt NES function in other proteins^{20–22}. The *rad24-nes* cells were shorter than wild-type and somewhat misshapen, appearing similar to

Δ *rad24* mutants⁹. Unlike wild-type Rad24, Rad24-*nes* was not excluded from the nucleus (Fig. 5a): ~0% (0 of 27) of the *rad24-nes* cells exhibited nuclear exclusion of Rad24-*nes*. Moreover, ionizing radiation failed to induce nuclear depletion of Cdc25 in *rad24-nes* cells (Fig. 5b). Nuclear accumulation of Cdc25 was detected in ~85% (11 of 13) of septated *rad24-nes* cells. Immunoprecipitation studies indicated that Cdc25 and Rad24-*nes* associate *in vivo* (data not shown). Thus, the defect in ionizing-radiation-induced nuclear depletion of Cdc25 in *rad24-nes* cells correlated with a defect in nuclear export of Rad24-*nes* protein.

In comparison with wild-type cells, the *rad24-nes* cells exhibited enhanced sensitivity ultraviolet irradiation, at a level similar to the sensitivity of Δ *rad24* mutants (Fig. 5c). We studied DNA-damage checkpoints by exposing synchronized cultures to the drug bleomycin, a radiomimetic drug that causes double-strand-DNA breaks^{23,24}. Continuous exposure to bleomycin prevented mitosis in >90% of wild-type cells (Fig. 5d). The bleomycin-induced checkpoint arrest was largely abolished in Δ *rad24* cells (Fig. 5e), as observed in previous studies in which DNA was damaged with ionizing radiation⁹. Likewise, the bleomycin-treated *rad24-nes* cells underwent mitosis with only slightly delayed kinetics relative to the mock-treated control culture (Fig. 5f). These findings indicate that an intact NES is important for the checkpoint function of Rad24.

Our results suggest the following model for control of the DNA-damage checkpoint. In undamaged cells, most Cdc25 is bound to

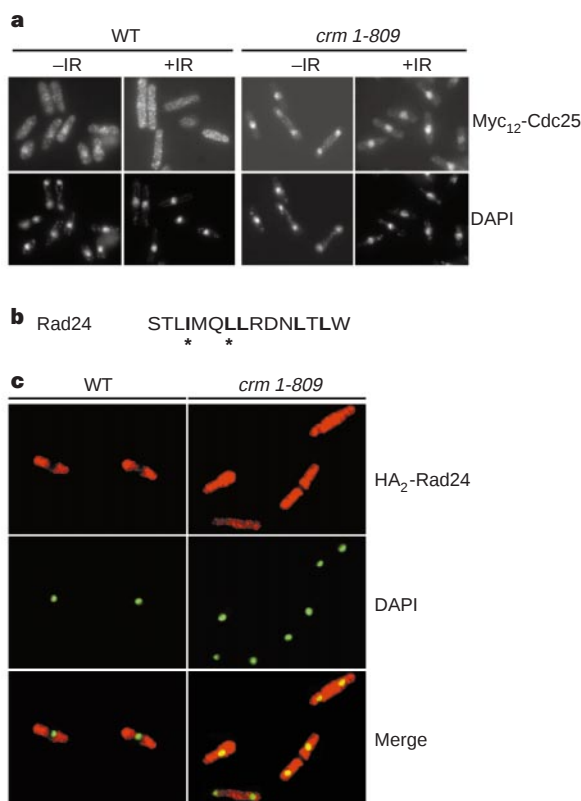


Figure 4 Crm1 is required for nuclear export of Cdc25 and Rad24. **a**, Nuclear localization of Cdc25 is enhanced in *crm1-809* cells. OM1730 (*cdc25:12myc*) and AL2277 (*cdc25:12myc crm1-809*) cells were grown at 30 °C and irradiated or mock-irradiated. Note that nuclear Myc₁₂-Cdc25 is detected in binucleate *crm1-809* cells, but not in binucleate wild-type cells. Nuclei were stained with DAPI. **b**, Residues 221–232 of Rad24 contain this leucine-rich NES-like sequence. Isoleucine and leucines are shown in bold. Asterisks indicate residues mutated to alanine in *rad24-nes*. **c**, Nuclear exclusion of Rad24 is disrupted in *crm1-809* cells. In the pseudocolour merged image, HA₂-Rad24 is red and nuclei are green. Regions of overlap are yellow.

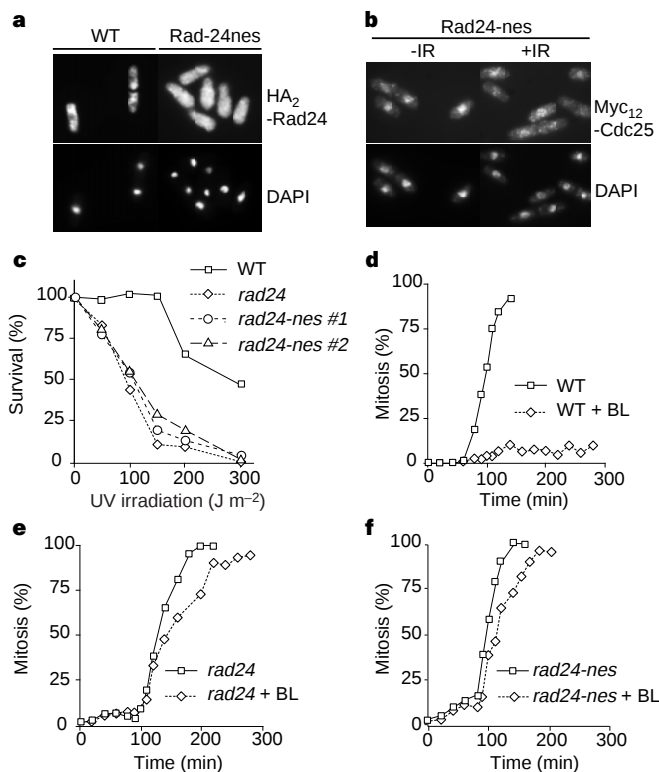


Figure 5 Analysis of *rad24-nes* mutants. **a**, Nuclear exclusion of Rad24-*nes* is impaired relative to wild-type (WT) Rad24. **b**, Irradiation-induced nuclear depletion of Cdc25 is defective in *rad24-nes* cells. **c**, Δ *rad24* and *rad24-nes* cells are similarly sensitive to ultraviolet (UV) irradiation. Strains BF1965 (WT), NR1589 (Δ *rad24*) and AL2281 (*rad24nes:2HA*; two different isolates) were assayed for survival after UV irradiation. **d–f**, *rad24-nes* cells are checkpoint-defective. Synchronous cultures of WT (**d**), Δ *rad24* (**e**) and *rad24-nes* (**f**) cells in early G2 phase were treated with bleomycin (BL) or mock-treated. Bleomycin prevented mitosis in wild-type cells, whereas Δ *rad24* and *rad24-nes* cells were checkpoint-defective.

Rad24 and excluded from the nucleus by a Chk1-independent process. In response to DNA damage, Chk1 phosphorylates the nuclear fraction of Cdc25 that is required to activate Cdc2. This phosphorylation induces binding of the nuclear fraction of Cdc25 to Rad24, a small protein that passively diffuses into the nucleus. Rad24 escorts the associated Cdc25 from the nucleus by a Crm1-dependent process; thus, in a sense, Rad24 functions as an attachable NES. It is worth noting that interpretation of phenotypes caused by *rad24* or *chk1* mutations is complicated by the presence of other kinases that phosphorylate Cdc25 and other 14-3-3 proteins that might associate with Cdc25^{9,14}. Our findings indicate that Chk1 may regulate the localization of Cdc25, but they do not exclude models in which Chk1 phosphorylation also directly inhibits Cdc25. However, *in vivo*, maintenance of this phosphorylation might be influenced by Rad24. 14-3-3 proteins have been proposed to protect phosphorylated sites from phosphatases^{25,26}. Thus, Rad24 might enhance the effect of Chk1 by isolating Cdc25 from nuclear phosphatases and shielding it from cytosolic phosphatases. □

Methods

Yeast strains and general methods. We used general genetic and biochemical methods for the study of fission yeast²⁷. Methods for expression, purification and immunoblot analysis of GST-fusion proteins²⁸, Cdc25 immunoprecipitation¹¹ and cell fractionation²⁹ have been described. Immunoreactive bands were revealed with horseradish-peroxidase-conjugated secondary antibodies and the SuperSignal western blotting detection system (Pierce). Cells were synchronized by centrifugal elutriation with a Beckman JE-5.0 rotor. Cells were irradiated with a ¹³⁷Cs source at 3 Gy min⁻¹ for 33 min to a total of 100 Gy. Ultraviolet irradiation was done with a Stratalinker UV 1800. Cells were plated at a density of 1 × 10³ cells per plate and irradiated with ultraviolet light at the indicated doses. Colonies were scored after 4 days' growth at 30 °C. Bleomycin sulphate (Sigma) was used at a concentration of 0.005 U ml⁻¹. Checkpoint studies were done with synchronous cultures made by centrifugal elutriation as described³, except that cells were briefly sonicated before elutriation to break apart clumped clusters. Cells were fractionated by digesting them with zymolyase followed by homogenization and centrifugation on linear Percoll gradients²⁹. Fractions were collected and analysed by phase-contrast and 4,6-diamidino-2-phenylindole (DAP)-stained fluorescence microscopy. Soluble components and small organelles remained on the top of the gradient (cytosolic), whereas nuclei were in fractions 24–27 (nuclear).

Indirect immunofluorescence microscopy. In the Cdc25- and Chk1-localization studies, cells were fixed with a 3.7% formaldehyde solution for 1 h at 30 °C. In the Rad24-localization studies, cells were collected by filtration on glass-fibre filters and fixed in methanol at -80 °C. After fixation, cells were washed three times in PEM buffer (100 mM PIPES, 1 mM EGTA, 1 mM MgSO₄, pH 6.9). The cell wall was digested at 37 °C for 10 min with 0.5 mg ml⁻¹ zymolyase 100T (Seikagaku America) in PEMS buffer (PEM buffer supplemented with 1 M sorbitol), followed by permeabilization with 1% Triton-X100. After washes, cells were blocked for 30 min at room temperature in PEMBAL buffer (PEM buffer, pH 6.9, supplemented with 1% BSA, 0.1% NaN₃, 100 mM L-lysine monohydrochloride). Primary antibodies (anti-Myc antibody (9E10; BaBCO), anti-haemagglutinin antibody (12AC5; BaBCO) and anti-GST antibody) were added to PEMBAL and incubated overnight at room temperature. After washing three times in PEMBAL, cells were incubated for 5 h at room temperature in PEMBAL containing Cy3-conjugated anti-mouse IgG or Cy3-conjugated anti-rabbit IgG (Jackson Laboratories) as secondary antibodies. The DNA stain DAPI was used to visualize nuclei. Cells were photographed using a Nikon Eclipse E800 microscope equipped with a Photometrics Quantix CCD camera. Images were acquired and quantified with IPLab Spectrum software (Signal Analytics Corporation). Cell-cycle positions were assigned as follows: early G2, uninucleate cells 7–8 µm in length; late G2, uninucleate cells 11–13 µm in length; M, binucleate cells with closely juxtaposed nuclei having condensed chromosomes; late M, binucleate cells with condensed chromosomes; G1/S, binucleate cells with interphase nuclei, often with septum.

In vitro mutagenesis and epitope tags. Codons corresponding to residues I222 and L226 of Rad24 were changed to alanine with the Altered Sites II *in vitro*

mutagenesis system (Promega). Plasmid pALTER-rad24 was constructed by inserting the 1.3-kilobase *Bam*HI fragment from pBF181 into pALTER-1. Plasmid pBF181 contains the *rad24⁺* open reading frame fused to a sequence encoding two copies of the haemagglutinin (HA) tag followed by six histidines³⁰. This plasmid was used as a template for the mutagenesis reaction. The primer sequence was 5'-CCTACAAAGATTCTACTTTAGCCATG-CAATTGGCGCGTGACAACC-3'. The mutated gene was cloned into pBF181, creating pAL183. Plasmids pBF181 and pAL183 were linearized using *Bgl*III and integrated at the *rad24⁺* locus in strain OM1730 (*h⁺cdc25:12myc:ura4⁺ura4-D18leu1-32*) or PR109 (*h⁻ura4-D18leu1-32*). These resultant strains expressed wild-type or mutant HA₂-Rad24 from the *rad24⁺* promoter. The *cdc25:6HA:ura4⁺* construct has been described⁴. The *cdc25:12myc:ura4⁺* construct was made using a similar strategy with tandem copies of the Myc tag.

Received 11 August; accepted 13 November 1998.

- Hartwell, L. H. & Weinert, T. A. Checkpoints: controls that ensure the order of cell cycle events. *Science* **246**, 629–634 (1989).
- Walworth, N., Davey, S. & Beach, D. Fission yeast chk1 protein kinase links the rad checkpoint pathway to cdc2. *Nature* **363**, 368–371 (1993).
- Rhind, N., Furnari, B. & Russell, P. Cdc2 tyrosine phosphorylation is required for the DNA damage checkpoint in fission yeast. *Genes Dev.* **11**, 504–511 (1997).
- Furnari, B., Rhind, N. & Russell, P. Cdc25 mitotic inducer targeted by Chk1 DNA damage checkpoint kinase. *Science* **277**, 1495–1497 (1997).
- Sanchez, Y. *et al.* Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497–1501 (1997).
- Peng, C. Y. *et al.* Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501–1505 (1997).
- Kumagai, A., Guo, Z., Emami, K. H., Wang, S. X. & Dunphy, W. G. The *Xenopus* Chk1 protein kinase mediates a caffeine-sensitive pathway of checkpoint control in cell-free extracts. *J. Cell Biol.* **142**, 1559–1569 (1998).
- Zeng, Y. *et al.* Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1. *Nature* **395**, 507–510 (1998).
- Ford, J. C. *et al.* 14-3-3 protein homologs required for the DNA damage checkpoint in fission yeast. *Science* **265**, 533–535 (1994).
- Boohar, R. N., Alfa, C. E., Hyams, J. S. & Beach, D. H. The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localization. *Cell* **58**, 485–497 (1989).
- Moreno, S., Nurse, P. & Russell, P. Regulation of mitosis by cyclic accumulation of p80^{cdc25} mitotic inducer in fission yeast. *Nature* **344**, 549–552 (1990).
- Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell* **84**, 889–897 (1996).
- Yaffe, M. B. *et al.* The structural basis for 14-3-3: phosphopeptide binding specificity. *Cell* **91**, 961–971 (1997).
- Peng, C. Y. *et al.* C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding. *Cell Growth Differ.* **9**, 197–208 (1998).
- Kumagai, A., Yakowec, P. S. & Dunphy, W. G. 14-3-3 proteins act as negative regulators of the mitotic inducer Cdc25 in *Xenopus* egg extracts. *Mol. Biol. Cell* **9**, 345–354 (1998).
- Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
- Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–310 (1997).
- Ossareh-Nazari, B., Bachelier, F. & Dargemont, C. Evidence for the role of Crm1 in signal-mediated nuclear protein export. *Science* **278**, 141–144 (1997).
- Stade, K., Ford, C., Guthrie, C. & Weiss, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041–1050 (1997).
- Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. & Luhrmann, R. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475–483 (1995).
- Fukuda, M., Gotoh, I., Gotoh, Y. & Nishida, E. Cytoplasmic localization of mitogen-activated protein kinase kinase directed by its NH2-terminal, leucine-rich short amino acid sequence, which acts as a nuclear export signal. *J. Biol. Chem.* **271**, 20024–20028 (1996).
- Gorlich, D. & Mattaj, I. W. Nucleocytoplasmic transport. *Science* **271**, 1513–1518 (1996).
- Suzuki, H., Nagai, K., Akutsu, E., Yamaki, H. & Tanaka, N. On the mechanism of action of bleomycin. Strand scission of DNA caused by bleomycin and its binding to DNA *in vitro*. *J. Antibiot.* **23**, 473–480 (1970).
- Kostrub, C. F., Knudsen, K., Subramani, S. & Enoch, T. Hus1p, a conserved fission yeast checkpoint protein, interacts with Rad1p and is phosphorylated in response to DNA damage. *EMBO J.* **17**, 2055–2066 (1998).
- Dent, P., Jelinek, T., Morrison, D. K., Weber, M. J. & Sturgill, T. W. Reversal of Raf-1 activation by purified and membrane-associated protein phosphatases. *Science* **268**, 1902–1906 (1995).
- Thorson, J. A. *et al.* 14-3-3 proteins are required for maintenance of Raf-1 phosphorylation and kinase activity. *Mol. Cell Biol.* **18**, 5229–5238 (1998).
- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Shiozaki, K. & Russell, P. Cell-cycle control linked to the extracellular environment by MAP kinase pathway in fission yeast. *Nature* **378**, 739–743 (1995).
- Hirano, T., Hiraoka, Y. & Yanagida, M. A temperature-sensitive mutation of the *Schizosaccharomyces pombe* gene *nuc2⁺* that encodes a nuclear scaffold-like protein blocks spindle elongation in mitotic anaphase. *J. Cell Biol.* **106**, 1171–1183 (1988).
- Shiozaki, K. & Russell, P. Stress-activated protein kinase pathway in cell cycle control of fission yeast. *Methods Enzymol.* **283**, 506–520 (1997).

Acknowledgements. We thank members of the cell-cycle labs at Scripps for support and encouragement, and M. Yanagida for *crm1-809* cells. A.L.-G. was supported by the MEC, Spain. This work was supported by a grant from the NIH (to P.R.).

Correspondence and requests for materials should be addressed to P.R. (e-mail: prussell@scripps.edu).

Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*

Richard A. Alm*, Lo-See L. Ling†, Donald T. Moir†, Benjamin L. King*, Eric D. Brown*, Peter C. Doig*, Douglas R. Smith†, Brian Noonan†, Braydon C. Guild†, Boudewijn L. deJonge*, Gilles Carmel†, Peter J. Tummino*, Anthony Caruso†, Maria Uria-Nickelsen*, Debra M. Mills†, Cameron Ives*, Rene Gibson†, David Merberg*, Scott D. Mills*, Qin Jiang‡, Diane E. Taylor‡, Gerald F. Vovis† & Trevor J. Trust*

* Astra Research Center Boston, 128 Sidney Street, Cambridge, Massachusetts 02139-4239, USA

† Genome Therapeutics Corporation, 100 Beaver Street, Waltham, Massachusetts 02453-8443, USA

‡ Department of Medical Microbiology & Immunology and Canadian Bacterial Diseases Network, University of Alberta, Edmonton T6G 2H7, Canada

Helicobacter pylori, one of the most common bacterial pathogens of humans, colonizes the gastric mucosa, where it appears to persist throughout the host's life unless the patient is treated. Colonization induces chronic gastric inflammation which can progress to a variety of diseases, ranging in severity from superficial gastritis and peptic ulcer to gastric cancer and mucosal-associated lymphoma¹. Strain-specific genetic diversity has been proposed to be involved in the organism's ability to cause different diseases or even be beneficial to the infected host^{2,3} and to participate in the lifelong chronicity of infection⁴. Here we compare the complete genomic sequences of two unrelated *H. pylori* isolates. This is, to our knowledge, the first such genomic comparison. *H. pylori* was believed to exhibit a large degree of genomic and allelic diversity, but we find that the overall genomic organization, gene order and predicted proteomes (sets of proteins encoded by the genomes) of the two strains are quite similar. Between 6 to 7% of the genes are specific to each strain, with almost half of these genes being clustered in a single hypervariable region.

H. pylori strain J99 (*cagA*⁺ *vacA*⁺), isolated in the USA in 1994 from a patient with a duodenal ulcer, was subjected to minimal subculturing before being sequenced by us in 1996. We describe this sequence below and compare it with the sequence of strain 26695, which was isolated in the UK before 1987 from a gastritis patient and which had a history of subculturing before being sequenced⁵. The J99 circular chromosome is 1,643,831 base pairs (bp) in size, which is 24,036 bp smaller than the 26695 chromosome. Several features, including the absence of an identifiable origin of replication, the average length of coding sequences and the relative frequency of the different initiation codons, are similar in the two strains (Table 1). We predict that there are 1,495 open reading frames (ORFs) in J99, representing 91% of the genome. Eighty-nine of these ORFs are absent from 26695. Of these J99-specific ORFs, 25 and 8 have sequence similarity to genes of predicted and unknown function, respectively, and 56 share no significant sequence similarity with any genes in public databases. J99 has 95 fewer genes than has been reported for 26695. However, 54 predicted genes of strain 26695 are less than 150 bp in size. In comparison with J99 genes, these 54 small genes either are highly conserved (16) and likely to encode proteins (note that three of these 26695 ORFs are part of larger ORFs in J99), or contain in-frame stop codons or exhibit nucleotide drift (38), as do other intergenic regions, and are therefore unlikely to encode

proteins. Thus, we revised the 26695 gene complement to 1,552 genes; 117 of these are unique to 26695 and 26 of these unique genes have a predicted function. Some genes appeared to contain a frameshift in J99 or 26695: 27 J99 genes are the equivalents of 55 predicted genes in 26695, and 7 genes from 26695 are the equivalents of 15 predicted genes in J99. In addition, three single-copy genes in 26695 have complete (gene HP1365; *H. pylori* 26695 genes are numbers preceded by 'HP') or partial (genes HP0818 and HP0928) duplications in J99. There are 1,406 genes in J99 that have counterparts in 26695.

Both genomes contain two 16S and two 23S–5S ribosomal RNA copies in the same relative locations, but strain 26695 contains a further, orphan 5S rRNA. In contrast to most other bacteria, the *H. pylori* rRNA loci are not contiguous, indicating that they may be regulated in a complex way. There are fewer complete insertion-sequence elements and fragments in J99 than in 26695, yet their location in both strains appears to be biased towards one half of the genome (Fig. 1a). Both genomes encode 36 transfer RNA species, each mapping to the same relative location. Neither strain contains Asn or Gln tRNA species; however, we have identified homologues for the *Bacillus subtilis* *gatABC* genes (gene JHP769/HP0830, JHP603/HP0658 and JHP909/HP0975; *H. pylori* J99 genes are numbers attached to the prefix 'JHP'), which amidate glutamate charged tRNAs to make glutamine-charged tRNAs⁶. Such genes are also likely to be responsible for amidation of appropriate aspartate-charged tRNAs.

Table 1 General comparative features of the *H. pylori* genomes

Genome features	<i>H. pylori</i> 26695	<i>H. pylori</i> J99
Size (base pairs)	1,667,867	1,643,831
(G + C) content (%)	39	39
Regions of different (G + C) content	8*	9†
AGTGATT repeats at bp = 1	26	2‡
<i>vacA</i> genotype	<i>slb/ml</i>	<i>sla/ml</i>
Open reading frames		
Per cent of genome (coding)	91.0	90.8
Predicted number	1,590	1,495
Functionally classified	875§	895
Conserved with no function	275§	290
<i>H. pylori</i> specific	345§	367
Number with signal sequence	517	502
Average length (base pairs)	954	998
Per cent AUG initiation codons	81.8	82.7
Per cent GUG initiation codons	9.7	6.7
Per cent UUG initiation codons	8.1	10.4
Per cent other initiation codons	0.4¶	0.2#
Insertion elements*		
Complete IS605 copies	5	0
Partial IS605 copies	8	5
Complete IS606 copies	2	1
Partial IS606 copies	2	2(4**)
RNA elements		
Per cent of genome (stable RNA)	0.75	0.75
23S–5S rRNA	2††(3‡‡)	2††
16S rRNA	2§§	2§§
tRNAs	36	36

* Includes the five reported previously⁵. Additional regions are HP0051–HP0054 and DNA flanking HP0611–HP0612 and HP0314–HP0316 (translocation 1).

† Four regions match those in 26695 (26695 loci 1 and 3 are joined in J99): JHP43–JHP46, JHP163–JHP165, JHP1422–JHP1423 and JHP414–JHP415 have a lower (G + C) content DNA flanking JHP299–JHP300 (translocation 1) has lower (G + C) content.

‡ Another cluster of these heptamer repeats is present ~2.35 kb upstream; at this position, there are 13 copies of the repeat in J99 and 2 copies in 26695.

§ Total ORFs equal 1,552 as defined during re-analysis of 26695 (see text).

¶ Defined as a *P* value of less than 0.05 using the SPScan algorithm in GCG 9.1.

†† HP0142 (CUG), HP0655 (AUU), HP0882 and HP0904 (AAA), HP0685 (GGA), and HP0451 (UGC).

JHP55, JHP402 and JHP600 (AUU).

* IS, insertion sequence.

** Two copies are smaller than the other two and are within the 31-bp repeated boundary of *cagPAI*.

†† 23S–5S rRNA is located at nucleotides 1,057,138–1,060,475 and 1,426,976–1,430,313 in J99, and 445,306–448,642 and 1,437,499–1,476,836 in 26695.

‡‡ 26695 orphan 5S rRNA is located at nucleotides 1,045,074–1,045,248.

§§ 16S rRNA is located at nucleotides 1,188,029–1,189,529 and 1,463,047–1,464,547 in J99, and 1,207,583–1,209,081 and 1,511,137–1,512,634 in 26695.

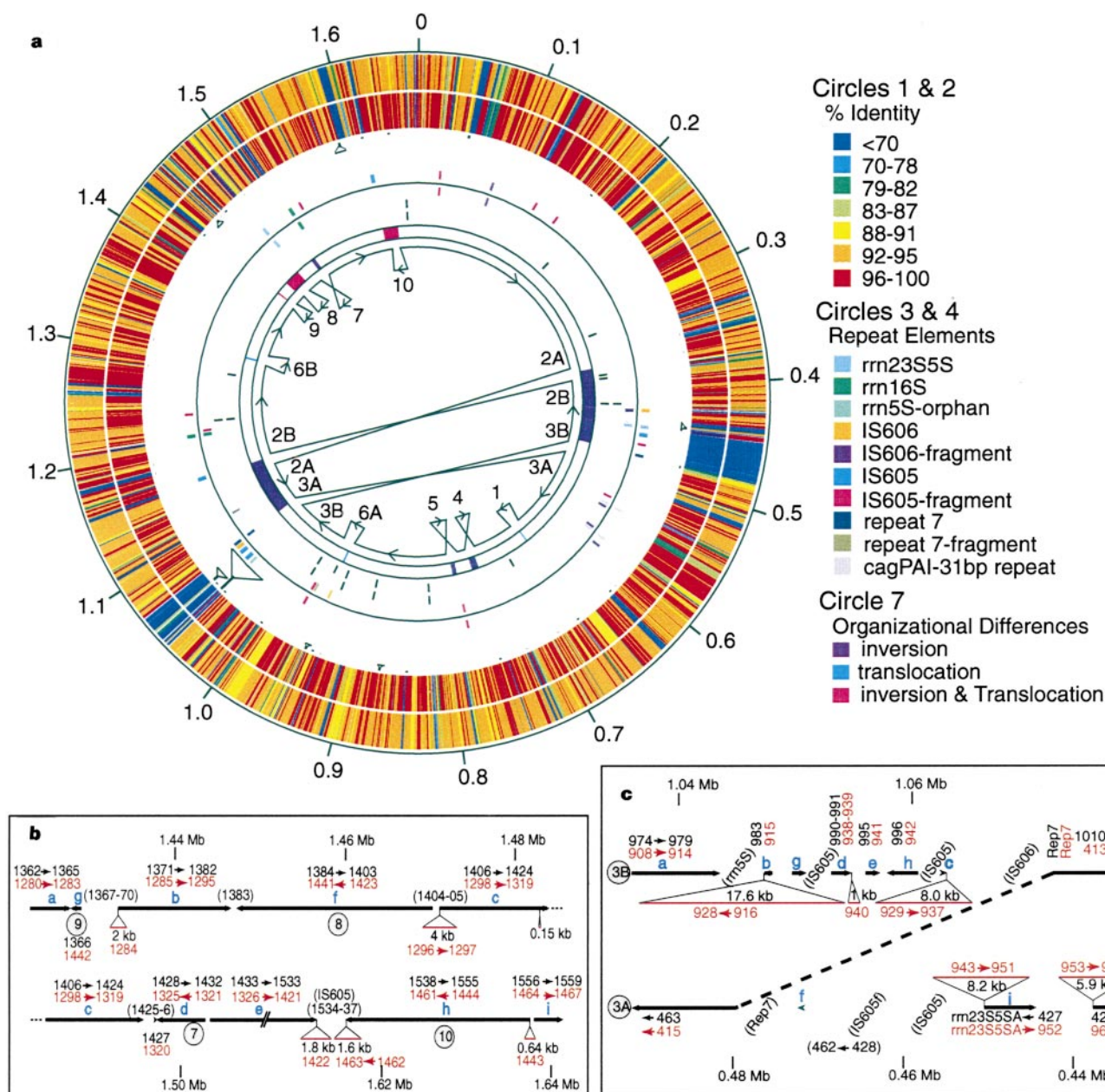


Figure 1 Comparison of the two sequenced *H. pylori* genomes based on the chromosomal organization of strain 26695. **a**, Genome-wide view. Circles are numbered starting from the outermost concentric ring. Circle 1, nucleotide and circle 2, amino-acid similarity between each J99 and 26695 orthologue. The relative location and amount of each J99-specific sequence are shown immediately inside the second circle (the height of each line is proportional to the amount of unique sequence, and for larger regions the size relative to the equivalent 26695 region is indicated by a triangle proportional to the 26695 scale). The largest J99-specific region shown is composed of two segments separated by 150 bp (see **b** for details). Circles 3 and 4, which flank the solid reference circle, show the locations of rRNA, insertion-sequence (IS) and repeat elements, for 26695 (circle 3) and J99 (circle 4). Circles 5 and 6 represent the locations of the *NotI* sites in the 26695 and J99 genomes, respectively. Circle 7 represents the relative transcriptional direction of J99 genes compared to their 26695 orthologues. Regions that are not coloured and translocations are transcribed in the same

relative direction in J99 and 26695, whereas inversions result in genes being transcribed in the opposite relative direction in J99. Circle 8 represents the organization of the J99 genome relative to the 26695 genome, incorporating artificial end points needed to allow the alignment. The required inversions and/or translocations are numbered consecutively for J99. **b, c**, An expanded view of the complex organizational differences 7–10 (**b**) and 3A/3B (**c**) shown in circle 8 of **a**. The 26695 ORFs are shown in the order and location that they are found (black numbers). The J99 ORFs are shown as red numbers. ORFs and other elements in parentheses are found in 26695 but not J99. The organization of J99 segments that share >90% identity to 26695 are depicted by the solid black lines, with arrows indicating the relative orientation of the J99 segments with respect to 26695 segments. The open triangles represent J99-specific DNA, drawn to scale, with the size and genes shown. The order of these regions in J99 is indicated by lower-case blue letters. The circled numbers correspond to the inversions/translocations referred to in **a**. Sizes of regions in **a** are shown in megabases (Mb).

Severity of *H. pylori* related disease is correlated with the presence of an island of genes (the *cag* pathogenicity island, *cagPAI*) associated with production of the CagA antigen⁷ and upregulation of interleukin (IL)-8 in gastric epithelial cells⁸. Both J99 and 26695 contain the complete *cagPAI* flanked by the same chromosomal genes and the previously described 31-bp repeat⁷ but lack the insertion-sequence 605 elements that are associated with *cagPAI* in strain NCTC11638 (ref. 7). Comparison of available *cagPAI* gene sequences showed minor differences between the J99 *cagPAI* genes and the other available sequences, such as apparent deletions in the *cag7* gene of J99 and 26695 (JHP476, HP0527) that lead to loss of up to 114 amino acids.

Like 26695, J99 encodes many families of paralogous proteins (337 genes, 22.5% of the total, are members of 113 families). One family contains the *vacA*-encoded vacuolating cytotoxin and three paralogues. Two of the three orthologues differ significantly in size between J99 and 26695: JHP856 encodes a protein that is 130 amino acids shorter than the protein encoded by HP922, and JHP556 represents a fusion between HP0610 and HP0609. The three paralogues in both strains lack the cleavage signal contained within *VacA* and may not be secreted.

The DNA-sequence differences between orthologues from the two strains are mainly found in the third position of coding triplets, consistent with the variance seen between *H. pylori* strains using methods dependent on the nucleotide sequence or on the sequencing of specific loci in different strains^{9–11}. However, this nucleotide variation does not translate into a highly divergent proteome (Fig. 1a). For example, there are only eight genes with $\geq 98\%$ nucleotide identity but 310 proteins with $\geq 98\%$ amino-acid conservation, including 41 with perfect identity.

To align homologous regions in the two genomes, we needed to artificially invert and/or transpose ten segments, ranging in size from 1 kilobase (kb) to 83 kb, of the J99 sequence. Most of the artificial end points are in intergenic regions and most are associated with insertion elements, repeated sequences or genes, and/or DNA-restriction/modification genes in one or both of the genomes (Fig. 1, Table 2), consistent with a possible role for such elements in generating these organizational differences. Two differences between the genomes are associated with genes encoding members of the large outer membrane protein (Omp) family. Inversion 5 in J99 could have resulted from a simple recombination across the inverted, repeated nucleotide sequence encoding the carboxy-terminal domain of two Omp proteins. Rearrangement 6 in J99 is the result of the equivalent of a reciprocal exchange of the Lewis-antigen-binding adhesin genes *babA* and *babB*¹²; BabA and BabB share similar C-terminal domains. The complex rearrangements

8–10 in J99 consist of both inversions and translocations (Fig. 1b). In both genomes, inversion 3 is associated with a region of (G + C) percentage that is lower (35%) than in the rest of the genome (39%). We named this region a 'plasticity zone' because it contains 46% and 48% of the genes that are unique to 26695 and J99, respectively (Fig. 1c). Although this region is continuous in J99, it is split in 26695 into two domains that are separated by ~ 600 kb. The presence of *vir* homologues, insertion sequences and a lower percentage of (G + C) DNA indicates that these regions might represent pathogenicity islands. The clustering of DNA with a lower (G + C) percentage is suggestive of horizontal DNA transfer, and the strain-specific sequence differences are consistent with different origins for this DNA. *H. pylori* and *Campylobacter* spp. plasmids have a (G + C) percentage in this lower range¹³. Significantly, two copies of the insertion-sequence 605 element and neighbouring 26695-strain-specific chromosomal DNA from the plasticity zone (genes HP0999–HP1001) are present on the *H. pylori* plasmid pHPM186 (GenBank accession number AF077006). Thus, plasmids may be responsible for the integration of new DNA into the *H. pylori* chromosome and for the transfer of this DNA between strains. Recombination across inverted repeats (repeat 7; ref. 5) in a progenitor strain that resembles strain 26695 would yield an arrangement similar to that of the region of inversion 3A in strain J99, but a similar reciprocal event cannot account for the complexity of the J99 3B locus (Fig. 1c).

To confirm the assembled sequence, we studied the J99 genome by pulsed-field gel electrophoresis (PFGE) and hybridization with specific probes (for J99 genes JHP117, 312, 548, 663, 733 and 1133–1136). Each observed *NotI* fragment was consistent in size with that predicted by the sequence. Hybridization of the 26695 genome *in silico* with these same probes yielded *NotI* fragments that were different in size to those observed with J99 DNA. Differences similar to those which we observed in restriction-fragment sizes and in probe hybridization patterns have been interpreted to mean that *H. pylori* strains are highly diverse in their genomic organization and gene order¹⁴. The differences in sizes of *NotI* fragments in strains J99 and 26695 are due mainly to silent nucleotide variation within genes. J99 contains twice as many *NotI* sites as 26695 (Fig. 1a), and silent nucleotide changes in 26695 are responsible for the absence of six of the seven *NotI* sites unique to J99; differences at the seventh site result in the alteration of a single amino acid. Similar minor sequence differences account for the variability in the *NruI*-site content between the strains. Thus, results obtained with lower-resolution techniques such as PFGE and polymerase chain reaction (PCR)-restriction-fragment length polymorphism (RFLP) have probably led to an overestimation of the true extent of genetic

Table 2 Elements associated with the artificial end points required to align the two *H. pylori* chromosomes

Locus	Type*	Size (kb)	Associated elements and genes	Strain
1	TR	1.5†	Lower (G + C)-content DNA (JHP299–JHP300; HP611–HP613); repeat element	Both
2A/B	IN	75	IS605 in 26695 (HP1095–HP1096); genes of unknown function in J99 (JHP331–JHP332) and 26695 (HP1094)	26695 or J99
3A/B	IN	83	'Orphan' 5S rRNA; inverted copies of repeat 7	26695
4	IN	10	Insertion of DNA-restriction/modification genes (JHP629–JHP630)	J99
5	IN	2.5	Conserved C terminus of <i>omp</i> genes (JHP659 and JHP662; HP0722 and HP0725); IS605 left-end fragment	Both
6A/B	TR	2†	Conserved C terminus of <i>bab</i> genes (JHP833 and JHP1164; HP0896 and HP1243) and 5' repeat element	Both
7	IN	5.5	Repeated, overlapping C terminus of histidine-rich genes (JHP1320–JHP1321; HP1427 and HP1432)	Both
8	IN/TR	24†	DNA-restriction/modification-gene replacement (JHP1296–JHP1297; HP1404/HP1405)	Both
10	IN/TR	21†	IS605 (HP1534–HP1535)	26695
9	IN/TR	1†	DNA-restriction/modification genes (JHP1442; HP1366); duplication of response regulator (JHP1283 and JHP1442; HP1365)	Both

* TR, translocation; IN, inversion.

† Distance between relative position of translocations in J99 and 26695 are 287 kb (locus 1), 385 kb (locus 6A/B), 146 kb (locus 8), 4 kb (locus 9) and 184 kb (locus 10).

diversity in *H. pylori*^{9,10,14}. However, these techniques will continue to be useful for epidemiology and strain discrimination.

To estimate the degree of conservation of gene order between J99 and 26695, we studied the immediate neighbours of each J99 gene and its 26695 orthologue, if present. Of the 1,495 genes in J99, 1,267 (84.7%) have the same neighbour on each side in both genomes; 161 (10.8%) are flanked by one common neighbour and one strain-specific gene; and 40 (2.7%) are flanked by strain-specific genes on both sides. Only 27 (1.8%) have the same neighbour on one side and a common gene that appears in a different position on the other side as the result of an organizational difference. There are 9 conserved gene strings that are more than 50 genes long, representing 46% of the genes common to both strains, with the longest string containing 133 genes. This highly conserved gene order indicates that physical linkage of a few genes (*topA/flaB*¹⁵ and *ftsH/pss/copA* (D.E.T., unpublished observations)) in several strains is the rule rather than the exception. The absence of extensive gene shuffling between J99 and 26695 is consistent with a low level of evolutionary divergence¹⁶.

Of the 1,495 J99 genes and the 1,552 re-annotated 26695 genes, 874 (58.5%) and 895 (57.7%) gene products, respectively, have been assigned putative functions. A total of 275 (18.4%) J99 and 290 (18.7%) 26695 gene products have orthologues of unknown function in other species, and 346 (23.1%) J99 and 367 (23.6%) 26695 genes are *H. pylori* specific (that is, they show no sequence similarity with genes available in public databases). Of these *H. pylori* specific

genes, 56 and 69 are specific to strains J99 and 26695, respectively. Excluding the strain-specific ORFs in the plasticity zone, the J99-specific genes are located singly (24 times) or as clusters of two (8 clusters) or three (1 cluster); many of these clusters appear to be organized to permit co-transcription. In one case, six genes (insertion-sequence 606 element and four J99-specific genes) are linked and are flanked by a duplicated region. In 17 corresponding locations, both J99 and 26695 have strain-specific genes. This high proportion of common relative loci for strain-specific ORFs indicates that *H. pylori* may have limited flexibility for containing strain-specific genes. Of the total of 206 strain-specific genes (89 in J99, 117 in 26695), the plasticity zones contain 94 (42 in J99, 52 in 26695); 125 of the 206 strain-specific genes (60.7%) are also specific to *H. pylori*, and 30 (14.6%) share similarity with genes of unknown function. J99- or 26695-specific genes have been assigned to the following categories: DNA restriction or modification (15 and 16, respectively); cell-envelope synthesis (4 and 2); cellular processes, such as DNA transfer and competence proteins (2 and 4); DNA replication (2 and 2); energy metabolism (2 and 1); and phospholipid metabolism (1 in 26695).

The fact that strain-specific DNA-restriction/modification genes have a lower (G + C) content than the remainder of the genome and are associated with regions that are organized differently in the J99 and 26695 genomes indicates that these genes may have been acquired horizontally from other bacterial species or transferred more recently from other *H. pylori* strains by natural transformation. Each *H. pylori* strain contains its own specific complement of these restriction/modification enzymes (R.A.A., unpublished observations). Nine type II methyltransferases are conserved between the two strains but lack identifiable cognate restriction-subunit partners, indicating that *H. pylori* may regulate gene expression by methylation.

The strain-specific genes encoding proteins involved in cell envelope (lipopolysaccharide and outer membrane protein) biosynthesis represent members of four paralogous families. Each strain contains one unique member of the *omp* families (HP0317 and JHP870). J99 and 26695 contain two (JHP820 and JHP1032) and one (HP1578) unique member, respectively, as well as four common members, of the *rfaI/rfaJ*-like family which is involved in lipopolysaccharide biosynthesis. In addition, J99 contains a unique member (JHP562) plus three common members of the *lex2B* lipopolysaccharide-biosynthesis family.

One of the J99-specific genes involved in energy metabolism encodes a second homologue of alcohol dehydrogenase (JHP1429), and the other (JHP585) may be required for amino-acid degradation. The 26695-specific energy-metabolism gene (HP1045) encodes an acetyl-CoA synthase. Strain 26695 has a second, larger acyl-carrier protein (encoded by HP0962) which is involved in phospholipid metabolism. J99 and 26695 have two (JHP919 and JHP931) and one (HP0440) unique genes encoding topoisomerase homologues, respectively, in their plasticity zones, which also contain the strain-specific genes that encode proteins involved in cellular processes. J99 has two adjacent *virB4* homologues (JHP917 and JHP918) which may have once represented a single complete gene, whereas 26695 contains two complete *virB4* (HP0441 and HP0459) and one truncated *virD4* (HP1006) homologues and a protein (encoded by HP0432) with similarity to a human protein kinase C.

The identification of homopolymeric tracts and dinucleotide repeats in *H. pylori* led to the prediction that 'slipped-strand repair' may modulate gene expression⁵, which could result in antigenic variation and adaptive evolution. The J99 gene sequences do not support some previously proposed examples of genes which are regulated in this fashion (for example, HP0211 and HP0928)¹⁷. In other cases, the data obtained from J99 do support this mechanism of control. Repeat lengths in some J99 genes differ from those in 26695 genes, indicating that such genes may be differently expressed in the two strains (Table 3). The same five members of

Table 3 Genes whose expression may be regulated by 'slipped-strand-repair'

JHP gene* (repeats:status)	HP gene* (repeats:status)	Repeat	Variation in J99 (#reads@#repeats)
Cell envelope (outer membrane protein)			
7 (6:off)	0009 (11:off)	(CT)	9@6
581 (9:on)	0638 (6:on)	(CT)	7@9
659 (9:off)	0722 (8:off)	(CT)	8@9; 1@8
662 (9:off)	0725 (6:off)	(CT)	7@9
1164 (8:on)	0896 (11:on)	(CT)	†
Cell envelope (lipopolysaccharide biosynthesis)			
86 (13:off)	0093-0094 (14:off)	(C)	18@13
194 (8:off)	0208 (11:off)	(AG)	†
563 (12:off)	0619 (13:off)	(C)	†
596 (5:on†)	0651 (13:on)	(C)	†
820 (14:on)	Absent§	(C)	†
1002 (13&9:off)	0379 (13&6†:on)	(C)&(A)	4@13&4@9
Cell envelope (flagella biosynthesis)			
625 (8:on)	0684-0685 (9:off)	(C)	†
Regulatory functions			
151 (9:on)	0164-0165 (13:off†)	(C)	3@8; 10@9; 3@10
Transport and binding proteins			
1129 (9:on†)	1206 (10:on)	(A)	6@9
DNA restriction/modification			
416 (10:off†)	0464 (15:on)	(C)	3@9; 3@10
1364 (14:on)	1471 (14:on)	(G)	†
1411 (11:off)	1522 (12:off)	(G)	†
1442 (8:off)	1366 (6:on)	(A)	1@7; 13@8; 5@9
Conserved with no known function			
131 (6:on)	0143 (7:off)	(A)	†
1312 (10:off)	1417 (9:off)	(G)	1@9; 3@10; 1@12
<i>H. pylori</i> specific with no known function			
203 (6&7:off)	0217 (12&6:on)	(G)&(G)	14@6&11@7; 1@6
351 (5:on)	1074 (6:off#)	(A)	12@5
681 (7:off)	0744 (9:off)	(AG)	†
1272 (13&12:off)	1353-1354 (12&15:off)	(C)&(C)	11@13&2@12; 2@13
1326 (11:on)	1433 (5:on†)	(C)	2@9; 8@10; 3@11; 4@12; 2@13
1392 (7:on)	1499 (6:off#)	(A)	14@7

* Gene number from J99 (JHP) or 26695 (HP). The number of repeats and whether the gene appears in-frame (on) or out-of-frame (off) are shown in parentheses.

† Insufficient sequence coverage was available to be deemed significant.

In addition to the repeats, an extra base pair of different identity to the repeat was found at one end.

§ Not applicable.

|| The string of C nucleotides in the *flp* gene (JHP625) has a C-to-A substitution in the middle.

¶ The 26695 gene also has another frameshift upstream of this string of C nucleotides.

The ORF predicted in 26695 (ref. 5) is truncated by this change in repeat length relative to J99.

the large *omp* paralogous family contain CT dinucleotide repeats in both strains, but the number of repeats differs without affecting the predicted expression status. The comparative data indicate that slipped-strand regulation may operate at two sites in some genes, including the α -(1,3)-fucosyltransferase gene. This regulatory mechanism also operates during laboratory passage of cell cultures: we found changes in the lengths of specific homopolymeric tracts or dinucleotide repeats within different populations of strain J99 (Table 3). We also detected nucleotide substitutions, most of which were found in the third position of coding triplets, at a low frequency.

Several factors could influence the pathophysiology and severity of disease associated with infection by different *cagA*⁺ *H. pylori* strains. First, strain-specific genes, such as those associated with the plasticity zone, could play a role. Second, differences in gene expression, perhaps mediated by slipped-strand repair, may be important and may affect the ability of the organism to colonize. Third, a human host factor(s) may play a significant, and perhaps unappreciated, part in susceptibility to, and severity of, *H. pylori* infection. In any host-parasite relationship, bacterial, host and environmental factors influence the host's susceptibility to and the clinical outcome of infection. For example, different mice strains exhibit markedly different susceptibilities to *H. pylori* colonization and clinical outcome¹⁸. Different human populations also show differences in susceptibility to *H. pylori* infection and incidence rates for gastric cancer¹⁹. Our identification of the minimal genetic diversity between two virulent strains, genes that are conserved between the two strains, and the strain-specific plasticity zone allows a better understanding of the biology of *H. pylori*. Our results suggest the need, and provide a unique opportunity, for a reassessment of the respective roles of bacterial and host factors in diseases associated with *H. pylori*. □

Methods

H. pylori strain J99 was sequenced, assembled and analysed nearly as described^{14,20}. The sequences of regions that differ significantly between strains J99 and 26695, including putative frameshifts, were all confirmed by sequencing PCR products of J99 and, where relevant, by diagnostic PCR of 26695. The nucleotide and amino-acid alignments used to determine the identity between orthologues were generated by ALIGN from version 2.0 of the FASTA program package. Paralogues were identified using BLASTP and TBLASTX algorithms.

The output was initially grouped such that all members of a family exhibited similarity to at least one other member, using a cut-off of $P < 10^{-10}$, and then checked manually for validity.

Received 4 September; accepted 24 November 1998.

1. Cover, T. L. & Blaser, M. J. *Helicobacter pylori* and gastroduodenal disease. *Annu. Rev. Med.* **42**, 135–145 (1992).
2. Atherton, J. C. Jr, Peek, R. M. Jr, Tham, K. T., Cover, T. L. & Blaser, M. J. Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. *Gastroenterology* **112**, 92–99 (1997).
3. Blaser, M. J. Not all *Helicobacter pylori* strains are created equal: should all be eliminated? *Lancet* **349**, 1020–1022 (1997).
4. Logan, R. P. H. & Berg, D. E. Genetic diversity of *Helicobacter pylori*. *Lancet* **348**, 1462–1463 (1996).
5. Tomb, J.-F. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547 (1997).
6. Curnow, A. W. et al. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94**, 11819–11826 (1997).
7. Akopyants, N. S. et al. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.* **28**, 37–53 (1998).
8. Censini, S. et al. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl Acad. Sci. USA* **93**, 14648–14653 (1996).
9. Akopyants, N., Bukanov, N. O., Westblom, T. U. & Berg, D. E. PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* **20**, 6221–6225 (1992).
10. Han, J., Yu, E., Lee, I. & Lee, Y. Diversity among clinical isolates of *Helicobacter pylori* in Korea. *Mol. Cells* **7**, 544–547 (1997).
11. Kansau, I. et al. Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. *Res. Microbiol.* **147**, 661–669 (1996).
12. Ilver, D. et al. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**, 373–377 (1998).
13. Lee, W. K. et al. Construction of a *Helicobacter pylori*–*Escherichia coli* shuttle vector for gene transfer in *Helicobacter pylori*. *Appl. Environ. Microbiol.* **63**, 4866–4871 (1997).
14. Jiang, Q., Hiratsuka, K. & Taylor, D. E. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Mol. Microbiol.* **20**, 833–842 (1996).
15. Suerbaum, S., Brauer-Steppkes, T., Labigne, A., Cameron, B. & Drlaca, K. Topoisomerase I of *Helicobacter pylori*: juxtaposition with a flagellin gene (*flaB*) and functional requirement of a fourth zinc finger motif. *Gene* **210**, 151–161 (1998).
16. Tatusov, R. L. et al. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* **6**, 279–291 (1996).
17. Saunders, N. J., Peden, J. F., Hood, D. W. & Moxon, E. R. Simple sequence repeats in the *Helicobacter pylori* genome. *Mol. Microbiol.* **27**, 1091–1098 (1998).
18. Sakagami, T. et al. Atrophic gastritis changes in both *Helicobacter felis* and *Helicobacter pylori* infected mice are host dependent and separate from antral gastritis. *Gut* **39**, 639–648 (1996).
19. Fock, K. M. et al. Seroprevalence of *Helicobacter pylori* infection. *Gastroenterology* **114**, A596 (1998).
20. Smith, D. R. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155 (1997).

Acknowledgements. We thank T. Dorman, M. Rubenfield, C. Butler, B. Andrews and K. MacCormack for assistance in finalizing and editing sequence information. D.E.T. is a member of the Canadian Bacterial Diseases Network and a Scientist with the Alberta Heritage Foundation for Medical Research.

Correspondence and requests for materials should be addressed to R.A.A. (e-mail: richard.alm@arcb.us.astro.com). The annotated genome sequence and further information are available in the *H. pylori* database (www.astro-boston.com/hpylori or www.genomecorp.com/hpylori) and will appear in GenBank under accession number AE001439.

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law.
Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Cubes, cameras and chromatography

The year's first crop of new products is eclectic as ever, including a clutch of HPLC columns, a trio of high-speed cameras and a pair of gloves.

Cubed Solutions

From Bio-Rad www.bio-rad.com

Buffers in a box: turn a tap for electrophoretic solutions.

Bio-Rad's latest range of electrophoresis buffers and solutions would not look out of place in the wine department of your local store. Called Cubed Solutions they are bulk-packed in 5-litre recyclable boxes fitted with a dispensing valve. They aim to both save on bench space and cut down on waste. Reagents so far available in this form are (all 10 ×) Tris/glycine buffer, Tris/glycine/SDS, Tris/borate/EDTA (TBE) and Tris/acetate/EDTA (TAE). Also available, this time at 20 ×, is SSC buffer.

Reader Enquiry No. 100

Paperwork online

Certificates of analysis

From Mallinkrodt Baker www.jtbaker.nl

Paper certificates of analysis could be history.

By keying-in the individual product code and lot number from the product label, Mallinkrodt Baker customers can obtain certificates of analysis including full analytical data and before-use information via the Internet. Also available from the company's website is a CD-Rom version of the catalogue.

Reader Enquiry No. 101

Certificates of analysis

From Sigma-Aldrich www.sigma-aldrich.com

A recently revamped website.

A new feature on the Sigma-Aldrich 'SuperSite' on the web is the round-the-clock availability of certificates of analysis covering the chemicals, biochemicals and immunochemicals available from the company. The analyses cover products from Sigma, Aldrich, Fluka and Supelco, giving analyses of each compound by lot number. Certificates can be printed straight from the screen and give detail on appearance, solubility and water contents, elemental analysis structure and, where applicable, potency.

Reader Service No. 102

Fashion accessories

Non-latex gloves

From Best www.bestglove.com

Thin but strong gloves suitable for sufferers of latex allergy.

Latex allergy and dermal sensitivity affect a significant number of laboratory workers



Handy in the lab: N-DEX gloves tear in extremis.

wearing disposable latex gloves for extended periods. N-DEX disposable gloves are an alternative. They contain no latex, as they are made with a low-modulus, soft-nitrile material just 0.1 mm thick. N-DEX gloves are available in both low-powder (non-talc, USP-grade vegetable starch) and powder-free styles. All meet US government CFR 21 specifications and are acceptable for use in food and pharmaceutical applications. The manufacturers claim that these non-latex gloves reduce hand fatigue caused by elastomeric stress and provide much greater puncture resistance than natural rubber and vinyl disposables of comparable thickness. In extreme cases where damage to the glove does occur, N-DEX gloves are designed to tear so the wearer knows that the glove should be replaced.

Reader Enquiry No. 103

Binocular telescopes

From Kent Scientific www.kentscientific.com

Making it easier to see the small things.

As research material keeps getting smaller, or is it the eyes getting older, thoughts may turn to the binocular telescopes made for Kent Scientific of Litchfield, California. They achieve 2 × magnification with a wide field of view and come fitted to a set of frames with a head band to help keep them on. Prescription lenses are available to special order.

Reader Enquiry No. 104



Double vision, courtesy of Kent Scientific.



Oriel Windows: high-speed images for the PC.

Images at speed

KSV high-speed cameras

From LOT-Oriel www.lot-oriel.com/uk

Digital cameras that have been used in automotive crash testing.

Available from LOT-Oriel, the KSV HiSiS 2000 series consists of two digital high-speed cameras. The HiSiS 2001 operates at 450 f.p.s. (frames per second) and the HiSiS 2002 at 2,200 f.p.s. High-speed images are converted to digital format and stored to an electronic memory or to a PC. Recordings are stored in standard Windows format. Available resolutions are 256 × 256, and 256 × 128 pixels. The real-time image is seen on the PC screen and the recordings are available for playback and processing. Computer requirement is for a 486 or Pentium in Windows 3.11 or WIN 95, with 16 MB RAM and 1 GB hard disk. Size is 168 × 118 × 69 mm and weight 1.8 kg.

Reader Enquiry No. 105

IES-1 digital cameras

From Xybion Electronic Systems

www.xybion.com

Lightening the darkness.

The IES-1 family of intensified digital cameras operates in the 400–900-nm spectral range, using a Gen III OMNI IV intensifier fibre-optically coupled to a 1,024 × 1,024-pixel Kodak Megaplug CCD progressive scan image array. Frame rate is 30 Hz, with 8-bit digital output as standard; 10-bit output and high-speed gating optional. Size is 8.51 × 3.58 × 3.1 inches and weight 4.5 pounds. Applications have included semiconductor inspection, day/night surveillance, process control and medical research.

Reader Enquiry No. 106

BioXight CCD cameras

From PixelVision Inc. www.pv-inc.com

Remote operation an option.

The BioXight series of high-performance CCD cameras are supplied by PixelVision Inc., of Beaverton, Oregon, with Windows 95/98 or NT-based PixelView image processing software. These cameras incorporate proprietary high-speed architecture for resolutions of up to $4,096 \times 4,096$ pixels and data collection at up to 80 Mbytes per second. Each camera is fitted with a multiple-input PCI computer interface claimed to allow error-free data transmission up to several kilometres distance, via fibre-optic cable.

Reader Enquiry No. 107

Apple Mac software

From Life Science Resources www.lsr.co.uk

Software to make Ultrapix cameras Mac-compatible.

New software developed by QED Imaging, and available from Life Science Resources, supports image capture from LSR's UltraPix range of high-speed digital cameras. The software allows for automatic correction of CCD effects, correction of uneven illumination and other adjustments to the output. QED is compatible with Power PCs and G3 Macs, operating on system 7.5 and above. LSR's digital cameras also interface to PCs running Windows via the UltraPlus software integral to the cameras.

Reader Enquiry No. 108

Hand-held technology

FieldSpec spectroradiometer

From Analytical Spectral Devices

www.asdi.com

A portable spectroradiometer, about the size of a pair of binoculars.

The FieldSpec HandHeld is a light-weight spectroradiometer with applications in remote sensing, agriculture and oceanography. It measures radiance, irradiance, CIE colour, colour temperature, luminance, illuminance, reflectance and transmittance. Running on standard camcorder batteries, the instrument is designed to be operated by the right or left hand and weighs about 2.7 pounds. The wavelength range is 300 to 1,100 nm, spectra resolution 3.5 nm, sampling interval 1.6 nm and it captures 10 spectra per second.

Reader Enquiry No. 109

Hygromer A1

From Rotronic ag www.rotronic.ch

Portable humidity monitor.

Rotronic's Hygromer A1 is a hand-held instrument designed for rapid and simple humidity measurement. It can check the air conditioning in the office or lab, and moni-



Just in case: Rotronic's probe is ultra-compact.

tor conditions for the storage of delicate equipment. Hygromer A1 features an adjustable probe position and a 'hold' function so that readings can be frozen and read later. Stability is claimed to be less than 1%rh per year. Either a rechargeable NiCd battery or 9-V alkaline block battery can be used.

Reader Enquiry No. 110

HPLC

HPLC and GLC training

from Savant Audiovisuals Inc.

A demo disk gives prospective clients a taste of Savant's training courses.

Eight programs for HPLC and GC training are sampled on a demo disk available from Savant. Seven programs relate to HPLC, including 'Introduction to HPLC', 'HPLC Calculations Assistant and Reference Tools' and 'Identification and Quantification Techniques in HPLC'. The eighth program is 'Introduction to Gas Chromatography'. The demo, and all the full training programs, run under Windows 3.1, 95, NT or 98.

Reader Enquiry No. 111

Oasis HLB column

From Waters Corporation www.waters.com

An HPLC column for ultrafast extraction of drugs from the blood.

The Waters Oasis HLB extraction column is designed for the ultrafast online analysis of drugs and their metabolites in plasma or serum. The extraction method, combining HPLC with tandem mass spectrometry, was developed by Glaxo Wellcome to measure 5–1,000 ng per ml quantities, faster than traditional methods and with the column able to take up to 300 injections of drugs before showing signs of deterioration. The Oasis HLB sorbent is a water-wettable copolymer available in cartridge, column and 96-well microtitre plate form. More

details are available in product bulletin No. WD162, available from Waters.

Reader Enquiry No. 112

Prostar chromatographs

From Varian

www.varian.com

HPLC customized to specific tasks.

Prostar modular liquid chromatographs can be configured as dedicated analytical HPLC systems, biocompatible HPLC instruments, or as high-throughput preparative units. They can be customized to an individual lab's specific requirements. For most research laboratories the Windows-based Star workstation is specified, offering single-point control, extensive data handling capacity and control of up to four instruments.

Reader Enquiry No. 113

Nitrogen detector

From Antek

www.antekhou.com

Chemiluminescence as a means of detecting nitrogen-containing compounds in HPLC.

The Model 8060 from Houston-based company Antek uses pyro-chemiluminescence to detect nitrogen-containing compounds. As each component elutes from the LC column it is combined with oxygen at a temperature of over 1,000 °C. The manufacturers claim that many of the obstacles to efficient nitrogen detection using conventional methods have been overcome, providing equimolarity for all nitrogen-bearing compounds, linearity over five orders of magnitude, and sensitivity to 0.1 g N.

Reader Enquiry No. 114

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

BACs

YACs

PACs

cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 30